

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁵ : C12P 21/00, 21/02, C12N 15/00 C12N 5/00	A1	(11) International Publication Number: WO 92/00382 (43) International Publication Date: 9 January 1992 (09.01.92)
(21) International Application Number: PCT/US91/04096 (22) International Filing Date: 14 June 1991 (14.06.91) (30) Priority data: <div style="display: flex; justify-content: space-between;"> <div>538,372</div> <div>15 June 1990 (15.06.90)</div> <div>US</div> </div> <div style="display: flex; justify-content: space-between;"> <div>614,452</div> <div>16 November 1990 (16.11.90)</div> <div>US</div> </div> (71) Applicant: CARNEGIE INSTITUTION OF WASHINGTON [US/US]; 1530 P Street, N.W., Washington, DC 20005 (US). (72) Inventor: LEE, Se-Jin ; 2509A Steele Road, Baltimore, MD 21209 (US). (74) Agents: KOKULIS, Paul, N. et al.; Cushman, Darby & Cushman, 1615 L Street, N.W., Eleventh Floor, Washington, DC 20036 (US).		(81) Designated States: AT (European patent), AU, BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent). Published <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
(54) Title: GDF-1 (57) Abstract <p>The present invention relates to a DNA segment encoding a mammalian GDF-1 protein and to the protein encoded therein. The invention further relates to a recombinant DNA molecule comprising a nucleotide sequence encoding mammalian GDF-1 protein, and host cells transformed therewith. The invention further relates to a mammalian UOG-1 protein and to a DNA segment encoding same.</p>		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinea	NL	Netherlands
BJ	Benin	GR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland	KR	Republic of Korea	SN	Senegal
CI	Côte d'Ivoire	LI	Liechtenstein	SU	Soviet Union
CM	Cameroon	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TC	Togo
DE	Germany	MC	Monaco	US	United States of America
DK	Denmark				

GDF-1BACKGROUND OF THE INVENTION5 1. Field of the Invention

 The present invention relates, in general,
to DNA segments encoding proteins of the
transforming growth factor beta superfamily. In
particular, the present invention relates to a DNA
10 segment encoding GDF-1, and unique fragments
thereof. The invention further relates to a
mammalian UOG-1 protein and to a DNA segment
encoding same.

15 2. Background Information

 A growing number of polypeptide factors
playing critical roles in regulating differentiation
processes during embryogenesis have been found to be
structurally homologous to transforming growth
factor β (TGF- β). Among these are Mullerian
20 inhibiting substance (MIS) [Cate et al, Cell 45:685-
698 (1986)], which causes regression of the
Mullerian duct during male sex differentiation; the
bone morphogenetic proteins (BMP's) [Wozney et al,
Science 242:1528-1534 (1988)], which can induce de
25 novo cartilage and bone formation; the inhibins and
activins [Mason et al, Nature 318:659-663 (1985);
Forage et al, Proc. Natl. Acad. Sci., USA 83:3091-
3095 (1986); Eto et al, Biochem Biophys Res Comm
142:1095-1103 (1987); and Murata et al, Proc. Natl.
30 Acad. Sci. USA 85:2434-2438 (1988)], which regulate
secretion of follicle-stimulating hormone by
pituitary cells and which, in the case of the
activins, can affect erythroid differentiation; the

Drosophila d capentaplegic (DPP) gene product [Padgett et al, Nature 325:81-84 (1987)], which influences dorsal-ventral specification as well as morphogenesis of the imaginal disks; the Xenopus Vg-1 gene product [Weeks et al, Cell 51:861-867 (1987)], which localizes to the vegetal pole of eggs; and Vgr-1 [Lyons et al, Proc. Natl. Acad. Sci., USA 86:4554-4558 (1989)], a gene identified on the basis of its homology to Vg-1 and shown to be expressed during mouse embryogenesis. In addition, one of the most potent mesoderm-inducing factors, XTC-MIF, also appears to be structurally related to TGF- β [Rosa et al, Science 239:783-785 (1988); and Smith et al, Development 103:591-600 (1988)]. The TGF- β 's themselves are capable of influencing a wide variety of differentiation processes, including adipogenesis, myogenesis, chondrogenesis, hematopoiesis, and epithelial cell differentiation [Massague, J., Cell 49:437-438 (1987)], and at least one TGF- β , namely TGF- β 2, is capable of inducing mesoderm formation in frog embryos [Rosa et al, Science 239:783-785 (1988)].

The present invention relates to a new member of the TGF- β superfamily, and to the nucleotide sequence encoding same. This new gene and the encoded protein, like other members of this superfamily, are likely play an important role in mediating developmental decisions related to cell differentiation.

30

SUMMARY OF THE INVENTION

It is a general object of the present invention to provide a novel cell differentiation regulatory factor and a nucleotide sequence encoding same.

In one embodiment, the present invention relates to a DNA segment encoding all, or a unique portion, of mammalian GDF-1, or a DNA fragment complementary to the DNA segment.

5 In another embodiment, the present invention relates to GDF-1 substantially free of proteins with which it is naturally non-covalently associated.

10 In a further embodiment, the present invention relates to a recombinantly or chemically produced GDF-1 protein having all, or a unique portion, of the amino acid sequence given in Figure 2, or functionally equivalent variations thereof.

15 In another embodiment, the present invention relates to a recombinant DNA molecule comprising the DNA segment of the present invention and a vector. The invention also relates to host cells stably transformed with the recombinant molecule.

20 Various other objects and advantages of the present invention will be apparent to one skilled in the art from the drawings and the description of the invention that follow.

BRIEF DESCRIPTION OF THE DRAWINGS

25 Figure 1 shows a Northern analysis of embryonic RNA. Two μ g of twice-poly A-selected mRNA isolated from day 8.5 post-coitum (p.c.) mouse embryos were electrophoresed on formaldehyde gels, transferred to nitrocellulose, and probed with GDF-
30 1 cDNA.

Figure 2 shows the sequence of GDF-1. The entire nucleotide sequence of GDF-1 derived from a single cDNA clone is shown with the predicted amino acid sequence below. The poly A tail is not shown.

Numbers indicate nucleotide position relative to the 5' end of the clone.

Figure 3 is a comparison of the predicted GDF-1 amino acid sequence with the amino acid sequences of previously-described members of the TGF- β superfamily.

(A): Alignment of the C-terminal amino acid sequences of GDF-1 (beginning at amino acid 236) with the corresponding regions of *Xenopus* Vg-1 [Weeks et al, Cell 51:861-867 (1987)], murine Vgr-1 [Lyons et al, Proc. Natl. Acad. Sci., USA, 86:4554-4558 (1989)], human BMP2a, 2b, and 3 [Wozney et al, Science 242, 1528-1534 (1988)] *Drosophila* DPP [Padgett et al, Nature 325:81-84 (1987)], human MIS [Cate et al, Cell 45:685-698 (1986)], human inhibin α , β A, and β B [Mason et al, Biochem Biophys Res Comm 135:957-964 (1986)], human TGF- β 1 [Derynck et al, Nature 316:701-705 (1985)], human TGF- β 2 [de Martin et al, EMBO J 6:3673-3677 (1987)], human TGF- β 3 [ten Dijke et al, Proc. Natl. Acad. Sci., USA 85:4715-4719 (1988); and Derynck et al, EMBO J 7:3737-3743 (1988)], chicken TGF- β 4 [Jakowlew et al, Mol Endocrinol 2:1186-1195 (1988)], and *Xenopus* TGF- β 5 [Kondaiah et al, J. Biol. Chem. 265:1089-1093 (1990)]. The 7 invariant cysteines are shaded. Dashes denote gaps introduced in order to maximize the alignment.

(B): Amino acid homologies among the different members of the superfamily. Numbers represent percent identities between each pair calculated from the first conserved cysteine to the C-terminus.

(C): Homology between GDF-1 and Vg-1 upstream of the presumed dibasic cleavage site. Two different regions are shown. A single gap of one amino acid has been introduced into the Vg-1

sequence in order to maximize the alignment.
Numbers indicate amino acid positions in the
respective proteins.

Figure 4 shows a sodium dodecyl sulfate-
polyacrylamide gel electrophoresis (SDS-PAGE) of the
in vitro translation product of GDF-1. Anti-sense
(lane 1) or sense (lanes 2-13) RNA, transcribed and
capped in vitro, was translated with a rabbit
reticulocyte lysate in the presence of
[³⁵S]methionine with (lanes 3, 5, 7, 9, 11, and 13)
or without (lanes 1, 2, 4, 6, 8, and 12) added dog
pancreas microsomes. Lanes: 2 and 3, translation
products from a full-length GDF-1 template; 4 and 5,
translation products from a deletion template
lacking the putative signal sequence; 6 and 7, Endo-
H treated translation products from a full-length
GDF-1 template; 8 and 9, trypsin-treated translation
products from a full-length GDF-1 template; 10 and
11, trypsin-treated translation products from a
deletion template lacking the putative signal
sequence; 12 and 13, translation products from a
full-length GDF-1 template treated with trypsin in
the presence of Triton X-100. Equal amounts of
products prepared in a single translation reaction
were used for lanes 2, 6, 8, and 12, for lanes 3, 7,
9, and 13, for lanes 4 and 10, and for lanes 5 and
11. Numbers at left indicate sizes of molecular
weight standards. The 41K, 39.5K, and 38K positions
were calculated relative to the mobilities of these
standards.

Figure 5 shows a genomic Southern analysis
of GDF-1. Ten µg of genomic DNA isolated from CHO
cells (hamster), BNL cells (mouse), or BeWo cells
(human) were digested with Eco RI (E), Bam HI (B),
or Hind III (H), electrophoresed on a 1% agarose

gel, transferred to nitrocellulose, and probed with GDF-1. Numbers at left indicate sizes (kb) of standards.

5 Figure 6 shows Northern analysis of embryonic RNA. Two μ g of twice-poly A-selected mRNA isolated from mouse embryos at the indicated days of gestation were electrophoresed on formaldehyde gels, transferred to nitrocellulose, and probed with GDF-1 cDNA. The assignment of the sizes of the bands
10 was based on the mobilities of RNA standards.

Figure 7 shows expression of GDF-1 in mouse tissues. Five μ g of once-poly A-selected mRNA isolated from various mouse tissues were
15 electrophoresed on formaldehyde gels, transferred to nitrocellulose, and probed with GDF-1 cDNA. The assignment of the size of the band was based on the mobilities of RNA standards.

Figure 8 shows expression of GDF-1 in the central nervous system. Two μ g of twice poly A-
20 selected mRNA isolated from fetal, neonatal, and adult brains, and from adult spinal cord, cerebellum, and brain stem were electrophoresed on formaldehyde gels, transferred to nitrocellulose, and probed with GDF-1 cDNA. The assignment of the
25 size of the band was based on the mobilities of RNA standards.

Figure 9 shows expression of GDF-1 in bacteria. Portions of GDF-1 cDNA were cloned into the pET3 vector and transformed into BL21 (DE3)
30 cells. Total bacterial extracts were electrophoresed on 15% SDS polyacrylamide gels and stained with Coomassie blue. The numbers at top indicate the first/last amino acid of GDF-1

contained in each construct. Numbers at left indicate sizes of molecular weight standards. Arrows at right indicate the positions of the bands representing GDF-1.

5 Figure 10 shows a schematic representation
of clones isolated from brain cDNA libraries. (A)
oligo dT-primed and random hexanucleotide-primed
murine brain cDNA libraries were prepared in the
lambda ZAP II vector (Stratagene) using the RNase H
10 procedure [Okayama et al, Mol. Cell. Biol. 2:161
(1982); Gubler et al, Gene 25:263 (1983)] according
to the instructions provided by Stratagene and
Amersham, respectively. Two separate oligo dT-
15 primed libraries of 0.7 million (library 1) and 2
million (library 2) recombinant phage and a random-
primed library of 1.3 million (library 3)
recombinant phage were obtained from 2 μ g of twice
poly A-selected adult brain mRNA per library.
Library 1 was amplified once prior to screening,
20 whereas libraries 2 and 3 were screened unamplified.
Hybridizations were carried out in 1M NaCl, 50 mM
sodium phosphate (pH 6.5), 2 mM EDTA, 0.5% SDS, and
10 X Denhardt's at 65°C. The final wash was carried
out in 0.5 X SSC at 68°C. (B) human adult cerebellum
25 and human fetal brain (17 to 18 week abortus) cDNA
libraries were obtained from Stratagene.
Hybridizations were carried out as for Figure 10(A)
except that the final wash was carried out in 2 X
SSC at 65°C. Numbers above the scales represent kb.
30 The locations of the UOG-1 and GDF-1, open reading
frames are shown by the solid and stippled boxes,
respectively. All clones were oriented and aligned
by determining the sequences at both ends.

35 Figure 11 shows the nucleotide sequences
of murine and human cDNA's encoding UOG-1 and GDF-

1. DNA sequences of both strands of murine (A) and human (B) cDNA clones were determined with the dideoxy chain termination method [Sanger et al, Proc. Natl. Acad. Sci., USA 74:5463 (1977)] using the exonuclease III/S1 nuclease strategy [Henikoff, Gene 28:351 (1984)]. The specific clones sequenced to assemble the complete sequences shown are described in the Examples below. Numbers indicate nucleotide position relative to the 5' end. The predicted amino acid sequences of UOG-1 and GDF-1 are shown below.

Figure 12 shows the hydropathicity profile of mUOG-1. Average hydrophobicity values were calculated using the method of J. Kyte and R.F. Doolittle, J. Mol. Biol. 157:105 (1982). Positive numbers indicated increasing hydrophobicity.

Figure 13 shows the alignment of murine and human sequences. Amino acid alignment of mGDF-1 with hGDF-1 (A) or mUOG-1 with hUOG-1 (B) were carried out using the SEQHP local homology program. Numbers indicate amino acid number relative to the N-terminus of each protein. Dashes denote gaps introduced in order to maximize the alignment. The 7 invariant cysteines in the GDF-1 sequences are shaded. The predicted dibasic cleavage sites are boxed. The box at position 145 in the mGDF-1 sequence shows the alternative amino acids at this position for GDF-1a (cysteine) or GDF-1b (serine). (C) DIAGON plot of murine and human nucleotide sequences was carried out with a window of 20 and stringency of 14. The locations of the UOG-1 and GDF-1 open reading frames are shown by the solid and stippled boxes, respectively. Numbers indicate nucleotide position in thousands.

Figur 14 sh ws a genomic Southern analysis f GDF-1. Ten micrograms f genomic DNA isolated from BNL cells (murin) or BeWo cells (human) were digested with Hind III (H), Bam HI (B), or Eco RI (R), electrophoresed on 1% agarose gels, transferred to nitrocellulose, and probed with the entire murine or human GDF-1 coding sequences as described in the legend to Figure 10. Filters hybridized with probes from the homologous species were washed in 0.2 X SSC at 68°C, whereas the filter containing human DNA probed with mGDF-1 was washed in 2 X SSC at 68°C. Numbers at left indicate sizes of standards in kb.

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to a DNA segment encoding all (or a unique portion) of GDF-1, a member of the transforming growth factor β superfamily. The invention further relates to the encoded protein (or polypeptide) and allelic and species variations thereof. A "unique portion" as used herein consists of at least five (or six) amino acids or, correspondingly, at least 15 (or 18) nucleotides. The present invention further relates to a recombinant DNA molecule comprising the above DNA segment and to host cells transformed therewith.

In particular, the present invention relates to a DNA segment that encodes the entire amino acid sequence given in Figure 2 (the specific DNA segment given in Figure 2 being only one such example), or any unique portion thereof. DNA segments to which the invention relates also include those encoding substantially the same protein as shown in Figure 2, including, for example, allelic variations and functional equivalents of the amino

acid sequence of Figure 2. The invention further relates to DNA segments substantially identical to the sequence shown in Figure 2. A "substantially identical" sequence is one the complement of which hybridizes to the sequence of Figure 2 at 68°C and 1M NaCl and which remains bound when subjected to washing at 68°C with 0.1X saline/sodium citrate (SSC) (note: 20 x SSC = 3M sodium chloride/0.3 M sodium citrate). The invention also relates to nucleotide fragments complementary to such DNA segments. Unique portions of the DNA segment, or complementary fragments, can be used as probes for detecting the presence of respective complementary strands in DNA (or RNA) samples.

The present invention further relates to GDF-1 substantially free of proteins with which it is normally non-covalently associated, or a unique peptide fragment of that protein. One skilled in the art can purify the GDF-1 using standard methodologies for protein purification. The GDF-1 protein (or functionally equivalent variations thereof), or peptide fragments thereof, to which the invention relates also include those which have been chemically synthesized using known methods. One skilled in the art will appreciate that multiple copies of the GDF-1 gene may exist. Each of the encoded proteins will likely carry out functions similar to or identical to the protein of Figure 2. Therefore, the term GDF-1 applies to these forms as well.

GDF-1 has potential N-linked glycosylation sites. Accordingly, one skilled in the art, without undue experimentation, can modify, partially remove or completely remove, the natural glycosyl groups from the GDF-1 protein using standard methodologies. Therefore, the proteins and peptides of the present invention may be glycosylated or unglycosylated.

The present invention also relates to recombinantly produced GDF-1 having the amino acid sequence given in Figure 2 or an allelic, or a functional equivalent, variant thereof. The recombinantly produced protein may be unglycosylated or glycosylated (the glycosylation pattern may differ from that of the naturally occurring protein). The present invention further relates to recombinantly produced unique peptide fragments of GDF-1.

The present invention also relates to a recombinant DNA molecule and a host cell transformed therewith. Using standard methodologies, well known in the art, a recombinant DNA molecule comprising a vector and a DNA segment encoding GDF-1, or a unique portion thereof, can be constructed. Vectors suitable for use in the present invention include, but are not limited to, baculovirus-derived vectors for expression in insect cells [Pennock et al, Mol. Cell. Biol. 4:399-406 (1984)], the T7-based expression vector for expression in bacteria [Rosenberg et al, Gene 56:125-135 (1987)] and the pMSXND expression vector for expression in mammalian cells [Lee and Nathans, J. Biol. Chem. 263:3521-3527 (1988)]. The DNA segment can be present in the vector operably linked to regulatory elements, for example, a promoter (e.g., polyhedrin, T7 or metallothionein I (Mt-I) promoters). The recombinant DNA molecule is suitable for transforming prokaryotic or eukaryotic cells.

The recombinant DNA molecule of the invention can be introduced into appropriate host cells by one skilled in the art using methods well known in the art. Suitable host cells include prokaryotic cells, such as bacteria, lower eukaryotic cells, such as yeast, and higher

eukaryotic cells, such as mammalian cells and insect cells.

5 The proteins and unique peptides of the invention can be used as antigens to generate GDF-1 specific antibodies using methods known in the art. Therefore, the invention also relates to monoclonal and polyclonal GDF-1 specific-antibodies.

10 The TGF- β superfamily encompasses a group of proteins affecting a wide range of differentiation processes. The structural homology between GDF-1 and the known members of the TGF- β superfamily and the pattern of expression GDF-1 during embryogenesis indicate that GDF-1 is a new member of this family of growth and differentiation factors. Based on the known properties of the other members of the this superfamily, GDF-1 can be expected to possess biological properties of diagnostic and/or therapeutic benefit in a clinical setting.

20 For example, one potential use for GDF-1 as a diagnostic tool is as a specific marker for the presence of tumors arising from cell types that normally express GDF-1. The availability of such markers would be invaluable for identifying primary and metastatic neoplasms of unknown origin or for monitoring the response of an identified neoplasm to a particular therapeutic regimen. In this regard, one member of this superfamily, namely, inhibin, has been shown to be useful as a marker for certain ovarian tumors [Lappohn et al, N. Engl. J. Med. 321:790 (1989)].

35 A second potential diagnostic use for GDF-1 is as an indicator for the presence of developmental anomalies in prenatal screens for potential birth defects. For example, abnormally high serum or amniotic fluids levels of GDF-1 may indicate the presence of structural defects in the

d vel ping fetus. Indeed, an ther embry nic marker, namely, alpha fetopr tein, is currently used r utinely in prenatal scr ens for neural tube defects [Haddow and Macri, JAMA 242:515 (1979)].

5 Conversely, abnormally low levels of GDF-1 may indicate the presence of developmental anomalies directly related to the tissues normally expressing GDF-1.

10 A third potential diagnostic use for GDF-1 is in prenatal screens for genetic diseases that either directly correlate with the expression or function of GDF-1 or are closely linked to the GDF-1 gene. Other potential diagnostic uses will become evident upon further characterization of the
15 expression and function of GDF-1.

20 Potential uses for GDF-1 as a therapeutic tool are also suggested by the known biological activities of the other members of this superfamily. For example, since some of these proteins act as cell-specific growth inhibitors, one potential therapeutic use for GDF-1 is as an anti-cancer drug to inhibit the growth of tumors derived from cell types that are normally responsive to GDF-1. Indeed, one member of this superfamily, namely,
25 Mullerian inhibiting substance, has been shown to be cytotoxic for human ovarian and endometrial tumor cells either grown in culture [Donahoe et al, Science 205:913 (1979); Fuller et al, J. Clin. Endocrinol. Metab. 54:1051 (1982)] or when
30 transplanted into nude mice [Donahoe et al, Ann. Surg. 194:472 (1981); Fuller et al, Gynecol. Oncol. 22:135 (1984)].

35 Conversely, if GDF-1 functions as a growth-stimulatory factor for specific cell types, other potential therapeutic uses will be apparent. For example, one member of this superfamily, namely, activin, has been sh wn t function as a nerve cell

survival molecule [Schubert et al, Nature 344:868 (1990)]. If GDF-1 possesses a similar activity, as is indicated by its specific expression in the central nervous system (see below), GDF-1 will likely prove useful in vitro for maintaining neuronal cultures for eventual transplantation or in vivo for rescuing neurons following axonal injury or in disease states leading to neuronal degeneration. Alternatively, if the target cells for GDF-1 in the nervous system are the support cells, GDF-1 will likely prove to be of therapeutic benefit in the treatment of disease processes leading to demyelination.

Many of the members of this superfamily, including GDF-1, are also likely to be clinically useful for tissue repair and remodeling. For example, the remarkable capacity of the bone morphogenetic proteins to induce new bone growth [Urist et al, Science 220:680 (1983)] has suggested their utility for the treatment of bone defects caused by trauma, surgery, or degenerative diseases like osteoporosis. Indeed, the bone morphogenetic proteins have already been tested in vivo in the treatment of fractures and other skeletal defects [Glowacki et al, Lancet i:959 (1981); Ferguson et al, Clin. Orthoped. Relat. Res. 227:265 (1988); Johnson et al, Clin. Orthoped. Relat. Res. 230:257 (1988)].

A determination of the specific clinical settings in which GDF-1 will be used as a diagnostic or as a therapeutic tool await further characterization of the expression patterns and biological properties of GDF-1 both under normal physiological conditions and during disease states. Based on the wide diversity of settings in which other members of this superfamily may be used for clinical benefit, it is likely that GDF-1 and/or

antibodies directed against GDF-1, will also prove to be enormously powerful clinical tools. Potential uses for GDF-1 will almost certainly include but not be restricted to the types of clinical settings described above. Moreover, as methods for improving the delivery of drugs to specific tissues or to specific cells become available, other uses for molecules like GDF-1 will become evident.

The following non-limiting Examples are provided to aid in the understanding of the present invention. In addition, data presented in the Examples (see, particularly, Examples 7 and 8) make possible a comparison of murine and human sequences derived from brain cDNA clones. The comparison reveals high conservation of two non-overlapping open reading frames. While the downstream open reading frame encodes GDF-1, the upstream open reading frame encodes a protein, designated UOG-1, containing multiple putative membrane-spanning domains. The data indicate that this mRNA gives rise to two different proteins. The bi-cistronic organization of UOG-1 and GDF-1 is unusual for eucaryotic mRNA's. Polycistronic mRNA's in procaryotes, however, often encode proteins carrying out related biological functions. Accordingly, UOG-1 and GDF-1 may functionally interact. The presence of multiple putative membrane scanning domains in UOG-1 indicates it may be a receptor, perhaps for GDF-1.

30

EXAMPLES

The following technical comments relate to the specific Examples that follow:

Construction and screening of an 8.5 day embryonic cDNA library: All embryonic materials were obtained

from random matings of CD-1 mice (Charles River). Mice were maintained according to the NIH guidelines for care and maintenance of experimental animals. The day on which the vaginal plug was noted was designated as day 0.5 p.c. Embryos were dissected out from the uterus, freed of all extra-embryonic membranes, and frozen rapidly. Total RNA was prepared by homogenization in guanidinium thiocyanate buffer and centrifugation of the lysate through a cesium chloride cushion [Chirgwin et al, Biochemistry 18:5294-5299 (1979)]. Poly A-containing RNA was obtained by twice-selecting with oligo-dT cellulose [Aviv, H., Proc. Natl. Acad. Sci. USA 69:1408-1412 (1972)]. A cDNA library was constructed in the lambda ZAP II vector using the RNase H method [Okayama et al, Mol. Cell Biol. 2:161-170 (1982); and Gubler et al, Gene 25:263-269 (1983)] according to the instructions provided by Stratagene. Recombinant plaques (3.2 million) were obtained from 2 µg of starting RNA. The library was screened with the oligonucleotide 5'-GCAGCCACACTCCTCCACCACCATGTT-3' (corresponding to the amino acid sequence NMVVEECGC) which had been end-labeled using polynucleotide kinase. Hybridization was carried out in 6X SSC, 1X Denhardt's, 0.05% sodium pyrophosphate, 100 µg/ml yeast tRNA at 50°C. Filters were washed in 6X SSC, 0.05% sodium pyrophosphate at 60°C.

DNA sequencing and blot hybridizations: DNA sequencing of both strands was carried out with the dideoxy chain termination method [Sanger et al, Proc. Natl. Acad. Sci., USA 74:5463-5467 (1977)] using the exonuclease III/S1 nuclease strategy [Henikoff S., Gene 28:351-359 (1984)].

For Northern analysis, RNA was electrophoresed in formaldehyde gels [(Lehrach et

al, Bi chemistry 16:4743-4751 (1977); and Goldberg, D.A., Proc. Natl. Acad. Sci., USA 77:5794-5798 (1980)], transferred to nitr cellulose, and hybridized in 50% formamide, 5X SSC, 4X Denhardt's, 0.1% SDS, 0.1% sodium pyrophosphate, 100 µg/ml salmon DNA at 50°C. Filters were washed first in 2X SSC, 0.1% SDS, 0.1% sodium pyrophosphate, then in 0.1X SSC, 0.1% SDS at 50°C.

For Southern analysis, DNA was electrophoresed on 1% agarose gels, transferred to nitrocellulose, and hybridized in 1M NaCl, 50 mM sodium phosphate, pH 6.5, 2 mM EDTA, 0.5% SDS, 10X Denhardt's at 65°C. The final wash was carried out in 2X SSC at 68°C.

In vitro translation experiments: The full-length 1387 bp GDF-1 cDNA or a deletion mutant lacking the first 251 nucleotides was subcloned into the Bluescript vector (Stratagene), and sense or anti-sense RNA was transcribed in vitro from the T3 or T7 promoters [Golomb et al, J. Virol 21:743-752 (1977); and McAllister et al, Nucl. Acids Res. 8:4821-4837 (1980)] in the presence of cap analog, as described by Stratagene. In vitro translations were carried out by incubating 0.5 µg RNA, 17.5 µl rabbit reticulocyte lysate (Promega), 20 µM cold amino acid mixture (Promega), and 20 µCi [³⁵S]methionine (New England Nuclear) in the presence or absence of 10 equivalents of dog pancreas microsomes (Promega) for 60 minutes at 30°C. Endoglycosidase digestions were carried out by diluting the translation reaction 1:30 with 100 mM sodium acetate pH 5.5, 0.1% SDS, 17 mU/ml endoglycosidase H (Boehringer-Mannheim). Protease digestions were carried out by diluting the translation reaction 1:20 with PBS, 1 mg/ml trypsin (Boehringer-Mannheim) in the presence or absence of

0.1% Triton X-100. All digesti ns were carried out
for 3 hours at 37°C. Translation products were
analyzed by electroph resis on 10% SDS
polyacrylamide gels [Laemmli, U.K., Nature 227:680-
5 685 (1970)] followed by fluorography with Enhance
(New England Nuclear).

Example 1

Cloning and Nucleotide Sequence of GDG-1

To identify new members of the TGF- β
10 superfamily that may be important for mouse
embryogenesis, a cDNA library was constructed in
lambda Zap II using poly A-selected RNA from whole
embryos isolated at day 8.5 p.c. As indicated
above, the library was screened with
15 oligonucleotides selected on the basis of the
predicted amino acid sequences of conserved regions
among members of the superfamily. Among 600,000
recombinant phage screened, the oligonucleotide
hybridized to 3 clones. Sequence analysis revealed
20 that the 3 cDNA clones were likely to represent
mRNA's derived from the same gene, which was
designated GDF-1.

Northern analysis of day 8.5 embryonic RNA
using the GDF-1 probe detected a single predominant
25 mRNA species of approximately 1.4 kb in length
(Figure 1). Because the original 3 cDNA isolates
were all smaller than 1.4 kb, portions of the
longest clone were used to re-screen the cDNA
library to isolate a full-length clone. Hybridizing
30 recombinant phage were seen at a frequency of
approximately 1 per 200,000.

The entire nucleotide sequence of the
longest cDNA clone obtained encoding GDF-1 is shown
in Figure 2. The 1387 bp sequence contains a single

1 ng pen r ading frame beginning with an initiating
ATG at nucl tid 217 and potentially encoding a
protein 357 amino acids with a molecular weight of
38,600. Upstream of the putative initiating ATG are
5 two in-frame stop codons and no additional ATG's.
Nucleotides 1259 to 1285 show a 25/27 match with the
complement of the oligonucleotide selected for the
original screening. The 3' end of the clone does
not contain the canonical AAUAAA polyadenylation
10 signal. Sequence analysis at the 3' end of 4
independent cDNA clones (all differing at their 5'
ends) showed that 2 clones terminated at the same
nucleotide, and the other 2 clones terminated at a
site 7 nucleotides further downstream (these clones
15 contained an additional AAAAATT sequence at the 3'
end).

Two cDNA clones isolated during this
screening process showed slight variations in their
sequence from that shown in Figure 2. In a limited
20 segment from which the nucleotide sequence was
determined, these 2 clones each showed 2 nucleotide
changes, one resulting in a cysteine to serine
substitution at amino acid 145 and the second
representing a third position change that did not
25 alter the amino acid sequence. These differences
are unlikely to be cloning artifacts since they were
found in independently-isolated clones. These
changes may represent allelic differences or they
may indicate the presence of multiple GDF-1 genes.

30 The predicted amino acid sequence
identified GDF-1 as a new member of the TGF- β
superfamily. A comparison of the C-terminal 122
amino acids with those of the other members of this
family is shown in Figure 3a. The predicted GDF-1
35 sequence contains all of the invariant amino acids
present in the other family members, including the 7
cysteine residues with their characteristic spacing,

as well as many of the other highly conserved amino acids. In addition, like other family members, the C-terminal portion of the predicted GDF-1 polypeptide is preceded by a pair of arginine residues at positions 236-237, potentially representing a site for proteolytic processing.

Figure 3b shows a tabulation of the percentages of identical residues between GDF-1 and the other members of the TGF- β family in the region starting with the first conserved cysteine and extending to the C-terminus. GDF-1 is most homologous to Vg-1 (52%) and least homologous to inhibin- α (22%) and the TGF- β 's (26-30%). Two lines of reasoning indicate that GDF-1 is not the murine homolog of Vg-1. First, GDF-1 is less homologous to Vg-1 than are Vgr-1 (59%), BMP-2a(59%), and BMP-2b (57%). Second, GDF-1 does not show extensive homology with Vg-1 outside of the C-terminal portion, and it is known that other members of this family are highly conserved across species throughout the entire length of the protein [Cate et al, Cell 45:685-698 (1986); Mason et al, Nature 318:659-663 (1985); Forage et al, Proc. Natl. Acad. Sci., USA 83:3091-3095 (1986); Derynck et al, Nature 316:701-705 (1985); Mason et al, Biochem. Biophys. Res. Comm. 135:957-964 (1986); and Derynck et al, J. Biol. Chem. 261:4377-4379 (1986)]. However, GDF-1 and Vg-1 do share two regions of limited homology N-terminal to the presumed dibasic cleavage site, as shown in Figure 3c.

Example 2

In vitro translation of GDF-1 RNA

The predicted GDF-1 sequence is also noteworthy for the presence of a core of hydrophobic

amino acids at the N-terminus, potentially
r presenting a signal sequence, as well as f r the
presenc of a p t ntial N-glyc sylati n site at
amin acid 191. To d t rmine wheth r these
5 sequences are functional and to confirm that
translation initiates as predicted at the first ATG,
in vitro translation experiments were carried out
using a rabbit reticulocyte lysate.

As shown in Figure 4 (lane 2), translation
10 of full-length sense GDF-1 RNA, transcribed and
capped in vitro, resulted in a major protein species
with a molecular weight of 39.5K, which agreed well
with the predicted molecular weight of 38.6K for the
translation product initiating at the most upstream
15 ATG; no such band was seen with translation of anti-
sense GDF-1 RNA (lane 1).

Support for translation initiation at the
most upstream ATG came from a starting DNA template
containing a deletion at the 5' end extending past
20 the first ATG codon resulting in a slightly smaller
translation product (lane 4), indicating that
translation in this case had initiated at the next
ATG codon (nucleotide 305). When full-length GDF-1
RNA was translated in the presence of dog pancreas
25 microsomes, some of the translated product migrated
slower than the full-length product (lane 3). This
slower migrating species (41K) could be converted to
a 38K form by treatment with endoglycosidase H (lane
7), consistent with the 41K and 38K species
30 representing the glycosylated and deglycosylated
forms, respectively, of the GDF-1 protein lacking a
signal peptide. Furthermore, the 41K species
(unlike the unprocessed 39.5K species) was resistant
to treatment with trypsin in the absence (lane 9)
35 but not in the presence (lane 3) of detergent,
suggesting that the 41K species was protected from
cleavage by its presence within the microsomes.

In contrast, parallel experiments carried out with protein translated from a deletion template lacking the signal sequence showed no shift to a high molecular weight species in the presence of microsomes (lane 5) and no protection from cleavage by trypsin (lane 11). Taken together, these data indicate that GDF-1 is a secreted glycoprotein like many of the other members of this superfamily.

Example 3

10 Southern blot analysis

To determine whether GDF-1 is a single-copy gene, Southern blot analysis was carried out using mouse genomic DNA as described above. As shown in Figure 5, the GDF-1 probe detected a single predominant band in 3 different digests of mouse DNA. However, even at high stringency, additional weakly hybridizing bands were detected. These minor bands are not likely to represent the products of partial digestion because many of these bands were smaller than the predominant band, and the intensities of these minor bands relative to the major band could be enhanced by reducing the stringency of the washing conditions.

Southern analysis was also extended to DNA isolated from other species. Even at high stringency, the GDF-1 probe detected a single predominant band in both hamster and human DNA (see Figure 5), indicating that GDF-1 is highly conserved across species. Moreover, as was seen with mouse DNA, additional minor bands could be detected in both human and hamster DNA at relatively high stringency.

Example 4

Expression of GDF-1

To determine the temporal pattern of expression of GDF-1 during embryogenesis, Northern analysis was carried out using poly A-selected RNA prepared from whole embryos isolated at days 8.5, 9.5, 10.5, 12.5, 14.5, 16.5, and 18.5. The GDF-1 probe detected two mRNA species showing distinct expression patterns (Figure 6). One mRNA species, 1.4 kb in length, was detected in embryos at days 8.5 and 9.5 but not in later stage embryos. The second mRNA species, 3.0 kb in length, appeared at day 9.5 and persisted throughout embryogenesis. The 1.4 kb species is likely to correspond to the GDF-1 cDNA sequence shown in Figure 2 since only the 1.4 kb species could be detected in day 8.5 embryos.

Northern analysis was also carried out using poly A-selected RNA prepared from a variety of adult tissues. As shown in Figure 7, the GDF-1 probe detected a 3.0 kb mRNA species expressed almost exclusively in the brain. Significantly lower, though detectable levels, were seen in the adrenal gland, ovary, and oviduct. No band corresponding to 1.4 kb was detected in any of these adult tissues. To further analyze the expression of the 3.0 kb mRNA in the brain, poly A-selected RNA was prepared from brains isolated at various developmental stages as well as from various subcompartments of the adult central nervous system. As shown in Figure 8, the GDF-1 probe detected a 3.0 kb mRNA species in embryonic and neonatal brains with the levels gradually increasing during brain development. Moreover, the 3.0 kb mRNA was also present at high levels in the spinal cord, cerebellum, and brain stem, suggesting that the expression of the 3.0 kb species may be widespread

in the central nervous system. In contrast, the 1.4 kb mRNA species was not detected in any of these samples.

5 In summary, the GDF-1 probe identified two mRNA species displaying distinct expression patterns. The 1.4 kb species, which corresponds to the cDNA sequence shown in Figure 2, was detected in embryos at day 8.5 and day 9.5 but not in later stage embryos or in any of the adult tissues tested. 10 The 3.0 kb species appeared at day 9.5, persisted throughout embryonic development, and was present almost exclusively in the central nervous system of adult animals. The 3.0 kb and the 1.4 kb species may be derived from two different genes or they may 15 represent alternatively initiated or processed transcripts, both derived from the GDF-1 gene.

Example 5

Preparation of antisera directed against GDF-1

20 Antibodies directed against GDF-1 can be used to characterize GDF-1 at the protein level. For this purpose, various portions of the GDF-1 protein have been overproduced in bacteria (Figure 9) using the T7-based expression vectors provided by Dr. F.W. Studier. Because the GDF-1 precursor is 25 likely to be cleaved approximately 120 amino acids from the C-terminus, several of these overproduced proteins can be used as immunogens to obtain antibodies directed against the mature C-terminus fragment as well as against the presumed pro- 30 region. Specifically, the GDF-1 fragments spanning amino acids 13 to 217 (which are fully contained within the pro- region) or amino acids 254-357 (which are fully contained within the mature C-terminal fragment) as well as the overproduced

25

pr t in extending from amino acids 13 to 357, have been excised from preparative SDS p lyacrylamide gels and can be used to immunize rabbits. S ra obtained fr m these rabbits following each boost can
5 be tested by Western blot analysis [Burnette, Anal. Biochem. 112:195-203 (1981)] of extracts prepared from bacteria harboring the overproducing plasmids. This analysis can reveal whether antibodies have been produced that recognize the bacterially-
10 produced immunogen. The animals can be boosted until a significant positive response is achieved as determined by this assay. To determine whether these antibodies also recognize nondenatured GDF-1, sense RNA derived from the full-length cDNA can be
15 transcribed (from the T3 or T7 promoters of subclones in the Bluescript vector), capped, and translated in vitro in the presence of [³⁵S]methionine. The antisera can then be tested for the ability to immunoprecipitate these translation
20 products.

Example 6Purification of GDF-1 from mammalian cells

In order to obtain GDF-1 to assay for biological activity, the protein can be overproduced
25 using the cloned cDNA. Because the pro- regions of the members of this superfamily appear to be necessary for the proper assembly of the active disulfide-linked dimers [Gray and Mason, Science 247:1328-1330 (1990)], and because proper assembly
30 and cleavage may not occur in bacteria, a mammalian cell line overproducing GDF-1 can be constructed. For this purpose, GDF-1 can be expressed in Chinese hamster ovary cells using the pMSXND expression vect r [Lee and Nathans, J. Biol. Chem. 263:3521-

3527 (1988)]. This vector contains a Mt-I promoter, a unique Xh I cloning site, splice and polyadenylation signals derived from SV40, a selectable marker for G418, and the murine dihydrofolate reductase (dhfr) gene under the control of the SV40 early promoter. The GDF-1 cDNA, truncated at the Hind III site in the 3' untranslated region, has been cloned downstream of the Mt-I promoter. The resulting construct, linearized at the unique Pvu I site (to enrich for integration events in this non-essential region), was transfected into CHO cells using the calcium phosphate method [Frost and Williams, Virology 91:39-50 (1978); van der Eb and Graham, Methods Enzymol. 65:826-839 (1980)]. G418-resistant clones can be grown in the presence of methotrexate to select for cells that amplify the dhfr gene and, in the process, co-amplify the adjacent GDF-1 gene. This vector and amplification scheme has been used in the past to construct a cell line in which one milligram of the desired protein was produced in seven 150 cm² tissue culture flasks [Lee and Nathans, J. Biol. Chem 263:3521-3527 (1988)]. In addition, because CHO cells can be maintained in a totally protein-free medium [Hamilton and Ham, in Vitro 13:537-547 (1977)], the desired secreted protein represented 10% of the total protein in the medium. This vector has also been made available to numerous other investigators, who have also overproduced their desired proteins in this manner [for example, see Colosi et al, Mol. Endocrinol. 2:579-586 (1988)].

Based on the results of in vitro translation experiments and on the known properties of other family members, it is likely that GDF-1 protein will be secreted into the medium. This can be verified by demonstrating the presence of GDF-1

in the conditioned medium of the overproducing cells by Western analysis. It also seems likely that the full length GDF-1 protein will be cleaved to generate the mature C-terminal fragment; indeed, such processing has been observed in the case of BMP-2a similarly overproduced in CHO cells [Wang et al, Proc. Natl. Acad. Sci., USA 87:2220-2224 (1990)]. Whether cleavage of GDF-1 takes place in the overproducing cells can be assessed by looking (by Western analysis) for the presence of a protein of the predicted size for the C-terminal fragment that reacts with antibodies directed against the C-terminal region but not with antibodies directed against the pro-region.

The mature GDF-1 protein can be purified from the conditioned medium of the producing cell line using standard protein separation techniques. An appropriate purification scheme can be empirically determined taking advantage of the known physical properties of other family members. For example, some of these proteins are known to have a high affinity for heparin [Ling et al, Proc. Natl. Acad. Sci. USA 82:7217-7221 (1985); Wang et al, Proc. Natl. Acad. Sci., USA 87:2220-2224 (1990)]. The final scheme can include an ion exchange chromatography step, a gel filtration step, and a reverse phase HPLC step. Each step of the purification can be monitored by electrophoresing column fractions on SDS polyacrylamide gels and identifying GDF-1 containing fractions by Western analysis. The purity at each step can be assessed by silver-staining of total proteins [Morrissy, Anal. Bioch. 117:307 (1981)]. The purified protein can be subjected to N-terminal amino acid sequencing to verify that the purified protein is GDF-1 and to precisely localize the site of cleavage from the precursor.

Example 7Cloning and nucleotide sequence of the
3.0 kb GDF-1 transcript

To determine whether the 3.0 kb band
5 represents an alternate transcript derived from the
GDF-1 gene or a transcript derived from a different
gene homologous to GDF-1, several cDNA libraries
were constructed from poly A-selected adult mouse
brain mRNA and screened with the 1.4 kb GDF-1 probe.
10 From approximately one million recombinant phage
screened from each of two separate oligo-dT primed
cDNA libraries, a single clone (mBr-1) was isolated
that hybridized with the GDF-1 probe at high
stringency. Seven hybridizing clones (mBr-2 through
15 mBr-8) were obtained by screening 0.6 million
recombinant phage from a random-primed cDNA library.
An additional 0.7 million recombinant phage from a
random-primed cDNA library were screened with a
probe derived from the 5' portion of clone mBr-7 to
20 obtain clones mBr-9 through mBr-14. Based on
partial nucleotide sequence analysis of the ends of
the clones, these 14 could be aligned within a
region spanning 2.7 kb (Figure 10a). The complete
2.7 kb cDNA sequence, obtained by determining the
25 entire nucleotide sequence of clones mBr-1, mBr-2,
and mBr-7, is shown in Figure 11a. Sequence
comparison showed that the previously-reported 1.4
kb sequence was essentially fully contained within
the 2.7 kb sequence (from nucleotides 1311 to 2687)
30 Within this region, the two sequences show three
nucleotide differences. The sequence derived from
clones mBr-2 and mBr-7 contains a C in place of T at
position 1725, an A in place of T at position 1960,
and a G in place of A at position 1974 compared to

the sequence derived from a day 8.5 embryo cDNA clone. Although two of these differences represent third position changes that do not alter the predicted amino acid sequence, one of the differences changes the cysteine at codon 145 to a serine. For simplicity, the coding sequence corresponding to a cysteine at position 145 will be referred to as GDF-1a, and the sequence corresponding to a serine at position 145 will be referred to as GDF-1b. To determine whether the expression of GDF-1a and GDF-1b is specific for the respective tissues from which they were isolated, the nucleotide sequences of 5 independent clones isolated from a day 8.5 embryo cDNA library and 7 independent clones isolated from an adult brain cDNA library were determined in a limited region spanning the nucleotide positions at which GDF-1a and GDF-1b differ. The sequence analysis revealed that of the 5 embryonic clones, 3 corresponded to GDF-1a, and 2 corresponded to GDF-1b; of the 7 brain clones, 2 corresponded to GDF-1a, and 5 corresponded to GDF-1b. Hence, both GDF-1a and GDF-1b appear to be expressed both in day 8.5 embryos, where only the 1.4 kb mRNA species could be detected, and in the adult brain, where only the 3.0 kb mRNA species could be detected. GDF-1a and GDF-1b may represent allelic differences or two different genes.

Upstream of the GDF-1 coding region, the 2.7 kb sequence contained an additional 1310 bp not present in the 1.4 kb sequence, leaving a total of 1527 bp upstream of the initiating codon for GDF-1. Unexpectedly, within this upstream region was a second long open reading frame beginning with a putative initiating methionine codon at nucleotide 74, extending for 350 codons, and terminating 404 nucleotides upstream of the GDF-1 initiating ATG. For simplicity, this second open reading frame will

b hereafter referred to as mUOG-1 (upstream of GDF-1). Because of the presence of multiple stop codons in the region between mUOG-1 and mGDF-1, at least 4 frameshifts would be required to translate the two open reading frames as a single protein. A search of the NBRF and GenBank sequence databases with the predicted mUOG-1 amino acid sequence and with the entire upstream nucleotide sequence, respectively, revealed no significant homologies with known sequences. However, hydropathicity analysis of the predicted mUOG-1 amino acid sequence revealed multiple clusters (at least seven) of hydrophobic residues, reminiscent of membrane spanning domains (Figure 12). Particularly striking is the most distal of these clusters, which is immediately followed by a highly charged C-terminal region. Like certain other proteins with multiple membrane-spanning domains [for example, see Nathans et al, Cell 34:807 (1983); Dixon et al, Nature 321:75 (1986)], mUOG-1 does not contain an obvious N-terminal signal sequence.

Example 8

Isolation of the human GDF-1 gene

In order to carry out sequence comparisons to look for potentially significant conserved regions in the GDF-1 mRNA and protein sequences, cDNA's encoding human GDF-1 were isolated using the murine GDF-1 probe. Three hybridizing clones (hBr-1 through hBr-3) were isolated from screening 0.6 million recombinant phage from a human adult cerebellum cDNA library (oligo dT-primed), and five clones (hBr-4 through hBr-8) were isolated from screening 1.4 million recombinant phage from a human fetal brain cDNA library (oligo dT/random hexanucleotide-primed) (Figure 10b). Figure 11b shows the 2510 bp human

cdna sequence obtained by determining the entire nucleotide sequence of clones hBr-5 and the 5'-most 400 nucleotides of clones hBr-6, hBr-7, and hBr-8. The 3'-half of the sequence contains a long open reading frame beginning with an ATG codon at nucleotide 1347 and potentially encoding a protein of 372 amino acids with a molecular weight of 38,853. The predicted amino acid sequence shows significant similarity to murine GDF-1 (Figure 13a).

Like the murine GDF-1 sequence, the human sequence contains a pair of basic residues (R-R) at amino acids 252-253, which presumably represents a site for proteolytic processing. Following the predicted cleavage site, the sequence contains all of the invariant and most of the highly conserved amino acids characteristic of all members of the TGF- β superfamily including the seven cysteine residues. The murine GDF-1 sequence and the human sequence are 87% identical in the region beginning with the first conserved cysteine and extending to the C-terminus and 69% identical throughout the entire length of the protein. Because other members of the TGF- β superfamily show a much higher degree of sequence conservation across species [Cate et al, Cell 45:685 (1986); Mason et al, Nature 318:659 (1985); Forage et al, Proc. Natl. Acad. Sci., USA 83:3091 (1986); Derynck et al, Nature 316:701 (1985); Mason et al, Biochem. Biophys. Res. Commun. 135:957 (1986); Derynck et al, J. Biol. Chem. 261:4377 (1986); de Martin et al, EMBO J. 6:3673 (1987); ten Dijke et al, Proc. Natl. Acad. Sci., USA 85:4715 (1988); Derynck et al, EMBO J. 7:3737 (1988); Miller et al, Mol. Endocrinol. 3:1108 (1989); *ibid*, p. 1926; Dickinson et al, Genomics 6:505 (1990)], genomic Southern analysis was carried out to determine whether the murine and human sequences represent the same gene. As shown in Figure 14, both murine and

human probes derived from the GDF-1 open reading frame hybridized to the same pattern of bands in human DNA, verifying that the human gene is indeed the homolog of murine GDF-1.

5 Like the murine sequence, the human sequence also contains a second long open reading frame potentially encoding 350 amino acids in the region upstream of the GDF-1 coding sequence. An alignment of this upstream open reading frame (hUOG-1) with that present in the murine sequence showed that the upstream open reading frame is even more highly conserved than that for GDF-1 (Figure 13b), with the overall amino acid sequence identity between mUOG-1 and hUOG-1 being 81%. Although the open reading frames for both mUOG-1 and hUOG-1 extend upstream of the putative initiating methionine to the very 5' ends of the sequences, two lines of reasoning suggest that these may be the true initiation codons. First, multiple cDNA's primed by random hexanucleotides at various distances from the 3' end terminated very close to the 5' ends of both the murine and human sequences (Figure 10). Second, the murine and human nucleotide and amino acid sequences show much less conservation upstream of the putative initiation codon for UOG-1 than in the coding sequence itself (Figure 13c). In contrast to the high degree of conservation observed between mUOG-1 and hUOG-1 and hUOG-1 and between mGDF-1 and hGDF-1, the intervening spacer region and the putative 5' and 3' untranslated regions show much less similarity between the murine and human sequences. This selective conservation of the two open reading frames is most clearly evident in a DIAGON plot comparing the murine and human nucleotide sequences (Figure 13c). The two sequences begin to diverge in the intervening spacer region precisely after the stop codons for UOG-1 and in the 3' untranslated

region 2 nucleotides following the stop codons for GDF-1. Moreover, the intervening spacer region in the murine sequence is 401 nucleotides in length whereas the corresponding region in the human sequence is only 269 nucleotides in length. The conservation of the amino acid sequence of UOG-1 is also evident in the non-random pattern of nucleotide differences between the murine and human sequences spanning the UOG-1 open reading frames. Of the 209 nucleotide differences in this region, 57 represent first position differences, 29 represent second position differences, and 123 represent third position differences; of the 123 third position differences, 89 do not result in differences in the predicted amino acid sequence.

* * * * *

All publication mentioned hereinabove are hereby incorporated by reference.

While the foregoing invention has been described in some detail for purposes of clarity and understanding, it will be appreciated by one skilled in the art from a reading of this disclosure that various changes in form and detail can be made without departing from the true scope of the invention.

WHAT IS CLAIMED IS:

1. A DNA segment encoding a mammalian GDF-1 protein, or an epitope specific thereto, or a DNA fragment complementary to said DNA segment.
2. The DNA segment according to claim 1 wherein said GDF-1 protein has the sequence as defined in Figure 2, 11A or 11B.
3. The DNA segment according to claim 1 wherein said mammal is a mouse, hamster or human.
4. A mammalian GDF-1 protein substantially free of proteins with which it is naturally non-covalently associated, or an epitope specific thereto.
5. The protein according to claim 4 which is unglycosylated.
6. The protein according to claim 4 wherein said mammal is a mouse, hamster or human.
7. The protein according to claim 4 wherein said protein is chemically synthesized.
8. The protein according to claim 4 wherein said protein has a sequence as defined in Figure 2, 11A or 11B, or functionally equivalent variation thereof.
9. A recombinantly produced GDF-1 protein having the amino acid sequence given in Figure 2, 11A or 11B, or functionally equivalent variation thereof.

10. The protein according to claim 9 wherein said protein is unglycosylated.

11. A recombinant DNA molecule comprising:

- i) said DNA segment according to claim 1;
- and
- ii) a vector.

12. A host cell stably transformed with said recombinant DNA molecule according to claim 11.

13. The host cell according to claim 12 wherein said cell is a procaryotic cell.

14. The host cell according to claim 12 wherein said cell is a eucaryotic cell.

15. A method of producing a recombinant GDF-1 protein, or functionally equivalent variation thereof, comprising culturing said host cell according to claim 12 under conditions such that said segment is expressed and said GDF-1 protein thereby produced, and isolating said GDF-1 protein.

16. A DNA segment encoding a mammalian UOG-1 protein, or an epitope specific thereto, or a DNA fragment complementary to said DNA segment.

17. A mammalian UOG-1 protein substantially free of proteins with which it is naturally non-covalently associated, or an epitope specific thereto.

18. A recombinantly produced UOG-1 protein having the amino acid sequence given in

Figure 11A or 11B, or functionally equivalent variation thereof.

19. A recombinant DNA molecule comprising:

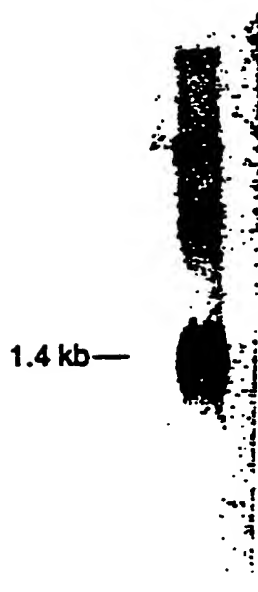
- i) said DNA segment according to claim 16; and
- ii) a vector.

20. A host cell stably transformed with said recombinant DNA molecule according to claim 19.

21. A method of producing a recombinant UOG-1 protein, or functionally equivalent variation thereof, comprising culturing said host cell according to claim 20 under conditions such that said segment is expressed and said UOG-1 protein thereby produced, and isolating said UOG-1 protein.

- 1 / 19

FIG. 1



- 2 / 19

1 CCCTTCTCCAGGGACTCTGGCTGCCAGCAGCTCCGCCTTTCAGATCAATTCTCGACCACC 60
61 CACCTTGGGACTGCGGCCCCAGTCCTGCCCTCTGGATCAGTGGGGTCCAGACACGCCCCCT 120
121 CCAGGACCTCAAAGCACCCCCGACCTAAGGTCACCAGCCCACTGGCCCCAGACGCAGTGG 180
181 GCTCCGCTGACTCTCTTGGACACCTCCTGGGAGGAAAATGCTCCCTGTCTGCCATCGTTT 240
M L P V C H R F
241 TTGCGACCACCTCCTCCTCTGCTCTTGTGCCCCGACGACCCTGGCCCCCGCGCCAGC 300
C D H L L L L L L L L P S T T L A P A P A
301 ATCCATGGGCCCCGCTGCGGCCCTGCTCCAGGTTCTTGGGCTTCCCGAAGCGCCCCGGAG 360
S M G P A A A L L Q V L G L P E A P R S
361 CGTCCCCACACACCGACCTGTGCCCTCCTGTCTGTGGCGCCTATTCCGTGCGCGTGACCC 420
V P T H R P V P P V M W R L F R R R D P
421 CCAGGAGGCCAGAGTGGGACGCCCCCTGTGCGGCCATGCCAGTGGAGGAACTAGGGGTCCG 480
Q E A R V G R P L R P C H V E E L G V A
481 CGGAAACATTGTGCGCCACATCCCCGACAGCGGTCTGTCTCCAGGCCCGACAACCCGC 540
G N I V R H I P D S G L S S R P A Q P A
541 CAGGACCTCGGGGCTGTGCCCCGAGTGGACAGTCGTCTTTGACCTGTGAATGTGGAGCC 600
R T S G L C P E W T V V F D L S N V E P
601 CACAGAGCGCCCAACACGCGCGCGCTTAGAGTTGCGGCTGGAGGCTGAGTGTGAAGATAC 660
T E R P T R A R L E L R L E A E C E D T
661 AGGAGGGTGGGAGCTAAGCGTGGCACTGTGGCCGACGCAGAGCATCCAGGGCCTGAGCT 720
G G W E L S V A L W A D A E H P G P E L
721 GCTGCGCGTGCCGGCGCCACCAGGGGTGCTCCTGCGCGCAGACCTACTGGGGACTGCAGT 780
L R V P A P P G V L L R A D L L G T A V
781 AGCCGCCAACGCATCAGTGCCCTGTACTGTGCGCCTGGCGCTGTCACTGCACCCTGGGGC 840
A A N A S V P C T V R L A L S L H P G A
841 CACTGCAGCCTGTGGGCGCCTGGCTGAGGCCTCCCTGTGCTGGTGACGCTGGACCCACG 900
T A A C G R L A E A S L L L V T L D P R
901 CCTGTGTCCCTTGCCGCGATTGCGCGCCACACGGAGCCAGGGTAGAAGTTGGTCCAGT 960
L C P L P R L R R H T E P R V E V G P V
961 GGGCACTTGTCTACCCGACGGTTGCATGTGAGCTTCCGTGAGGTGGGCTGGCACCGTTG 1020
G T C R T R R L H V S F R E V G W H R W
1021 GGTGATCGCGCCGCGTGGCTTCCTAGCCAACTTCTGCCAGGGCACGTGCCCACTACCCGA 1080
V I A P R G F L A N F C Q G T C A L P E
1081 AACGCTGAGGGGACCCGGCGGGCCGCTGCACTCAACCACGCJGTGCTGCGCGCGCTCAT 1140
T L R G P G G P P A L N H A V L R A L M
1141 GCACGCAGCTGCTCCCAACCCCGGGTGCAGGCTCGCCCTGCTGCGTGCCAGAGCGTCTATC 1200
H A A A P T P G A G S P C C V P E R L S
1201 ACCCATCTCCGTGCTCTTCTTCGACAATAGTGACAACGTGGTCCCTGCCACACTACGAAGA 1260
P I S V L F F D N S D N V V L R H Y E D
1261 CATGGTGGTGGATGAGTGTGGCTGCCGTTGACCACCCGGGACACCCTTTCAGGGACCGCC 1320
M V V D E C G C R
1321 CCACGCAAAAGCAGGGACTGTTTGTTCATGTTTTATTGGTGACAAAAAGCTTAAACAAA 1380
1381 TTTGACT 1387

- 3 / 19

Fig. 3a

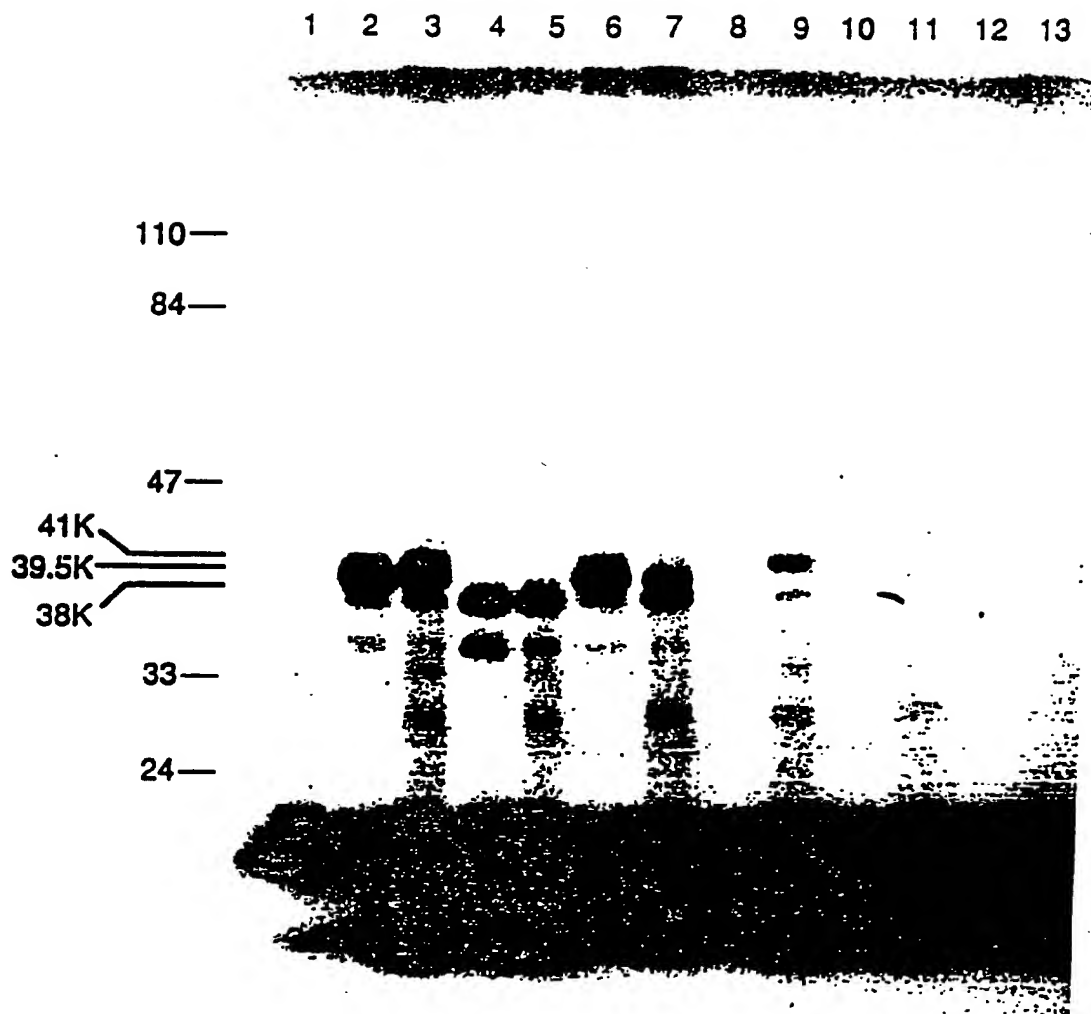
GDF-1 RRHTEPRVEVG--PVGTQTRRLRHVSE-REVGWHRWVIAPRGFLANF--CQGTALPETLRGPGGPP
 Vg-1 RKRSYSKLPET--ASNIKKRHLYVEE-KDVGWQNWVIAPOQGYMANYSYGE--P-YPLTEILNG--
 BMP-1 GSGSSDYNGSE--LKTAQKKHLYVSE-QDLGWQDWIIVAPRGYAANYDGE--P-SFPLNAHMNA--
 BMP-2a KQAKHKQRKR--LKSSQKRHPLYVDE-SDVGWQDWIIVAPRGYAANYDGE--P-SFPLADHLNS--
 BMP-2b TLKKHHSQRAR--KNKNQKRHSLYVDE-SDVGWQDWIIVAPRGYAANYDGE--P-FPLADHLNS--
 DPP -HARRRTRRKN--HDDTQARRSLKVDE-ADIGWSEWIIISPKSFDAYYCHG--P-QEPMFKSLKPS--
 MIS GFAQRSAGATA--ADGPAARRSLYVDE-SDVGWQDWIIVAPRGYAANYDGE--P-FPLADHLNS--
 Inhibin α RLLQRPPEPA--AHANQRRVALNISE-QELGWERWIVPPSYIEHYCHG--P-GMPQSDRNPY--
 Inhibin β GLE--CDGKV--NLCQKQOFFVSE-KDIGWQDWIIVAPRGYAANYDGE--P-GMPQSDRNPY--
 Inhibin β GLE--CDGRT--NLCQKQOFFVSE-KDIGWQDWIIVAPRGYAANYDGE--P-GMPQSDRNPY--
 TGF-1 ALDTNYCFSS--QDNCQVRQ--LYIDFKDLGK--WIHEPKGYHANF--P-YIW--SSD--
 TGF-2 ALDAAYCFRNV--QDNCQVRP--LYIDFKDLGK--WIHEPKGYHANF--P-YIW--SSD--
 TGF-3 ALDTNYCFRNV--QDNCQVRP--LYIDFKDLGK--WIHEPKGYHANF--P-YIW--SSD--
 TGF-4 DLDTDYCFGPGTDEKNCQVRP--LYIDFKDLGK--WIHEPKGYHANF--P-YIW--SSD--
 TGF-5 GVGQEYCFGNN--GPNCQVRP--LYIDFKDLGK--WIHEPKGYHANF--P-YIW--SSD--

GDF-1 ALNHAVLRLMHAAA-PTPGAGSECCV--PERLSPISVLFF-DNSDNVVLRLHYEDMVVDECCV
 Vg-1 -SNHAILQTLVHS--IEPEDIPLECCV--PTKMSPI SMLFY-DNNDNVVLRLHYENMAVDECCV
 Vgr-1 -TNHAIIVQTLVHL--MNPEYVPKECCV--PTKLNALISVLYF-DDNSNVILKRYRNMVVRACCH
 BMP-1 -TNHAIIVQTLVNS--VNSKIPKACCV--PTELSAISMLYL-DENEKVLKNYQDMVVEGCH
 BMP-2a -TNHAIIVQTLVNS--VNSKIPKACCV--PTELSAISMLYL-DENEKVLKNYQDMVVEGCH
 BMP-2b -TNHAIIVQTLVNS--VNSKIPKACCV--PTELSAISMLYL-DENEKVLKNYQDMVVEGCH
 DPP -TNHAIIVQTLVNS--VNSKIPKACCV--PTELSAISMLYL-DENEKVLKNYQDMVVEGCH
 MIS -TNHAIIVQTLVNS--VNSKIPKACCV--PTELSAISMLYL-DENEKVLKNYQDMVVEGCH
 Inhibin α -GNAPPT--KMQARGAALARPCCV--PTAYAGKLLISLSEER--ISAHVHPNMVATECCV
 Inhibin β -SEHSTVINHYMRGLNPGT-VNS--PTKLRPMSMLYF-DDGONIKKDIQNMVVEGCH
 Inhibin β -SEHSTVINHYMRGLNPGT-VNS--PTKLRPMSMLYF-DDGONIKKDIQNMVVEGCH
 TGF-1 -TOYSKVLALYN--QHNPASASAPCCV--POALEPLIIVY-DGRKPKV-EQLSNMIVRSCH
 TGF-2 -TOYSKVLALYN--QHNPASASAPCCV--POALEPLIIVY-DGRKPKV-EQLSNMIVRSCH
 TGF-3 -TOYSKVLALYN--QHNPASASAPCCV--POALEPLIIVY-DGRKPKV-EQLSNMIVRSCH
 TGF-4 -TOYSKVLALYN--QHNPASASAPCCV--POALEPLIIVY-DGRKPKV-EQLSNMIVRSCH
 TGF-5 -TOYSKVLALYN--QHNPASASAPCCV--POALEPLIIVY-DGRKPKV-EQLSNMIVRSCH

[illegible]

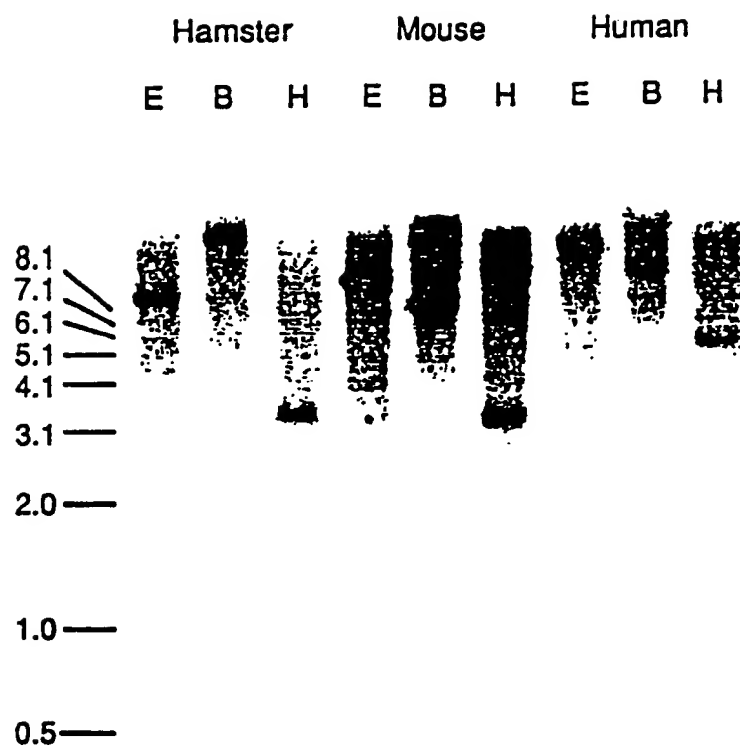
54	133
GDF-1	PVPEVMWRLFRRRDPQEARVGRPLRPCHVVEELGVAGNIVRHIPDSGLSSRPAQPARTSGLCPEWTVVFDLSNVEPTERT
Vg-1	PVPSILWRIFNQRMGSSIQKKKPDLCFVEEFNVPGSVIRVFPDQGRFIIPYSDDIHPTQCLEKRLFFNISAIEKEEVRT
46	124
219	234
GDF-1	SLLLVTLDPRLCPLPR
Vg-1	SLLTVTLNPLRCKRPR
228	243

FIG. 4



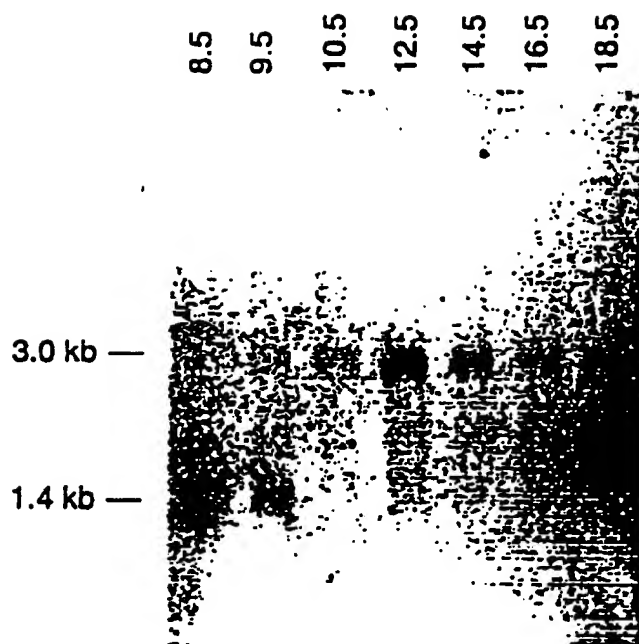
- 7 / 19

FIG. 5



- 8 / 19

FIG. 6



- 9 / 19

FIG. 7

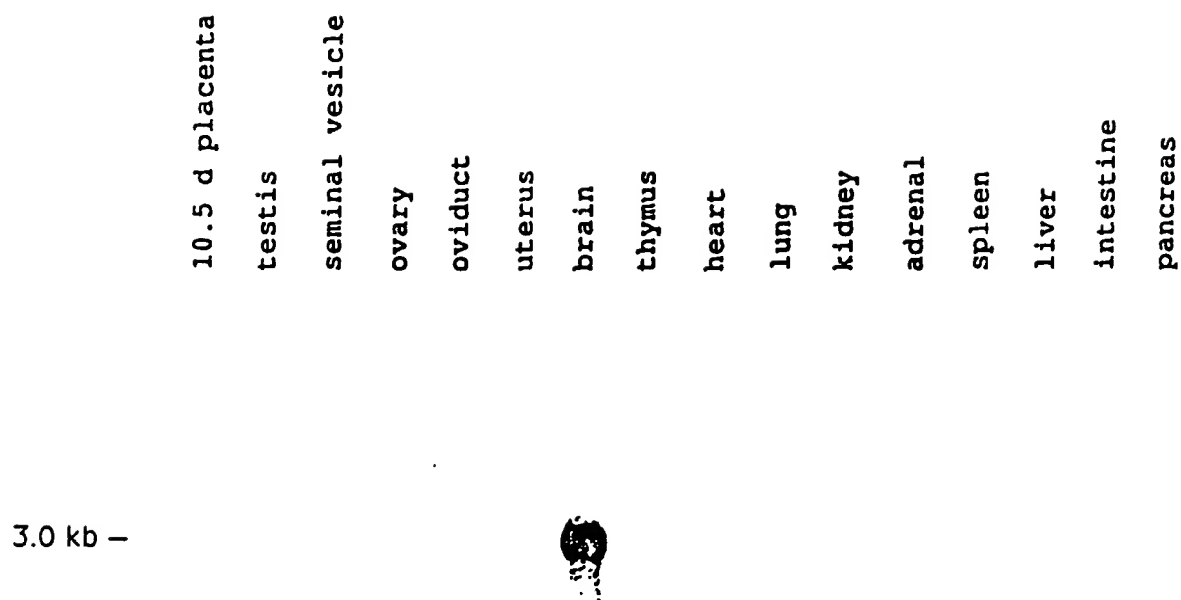
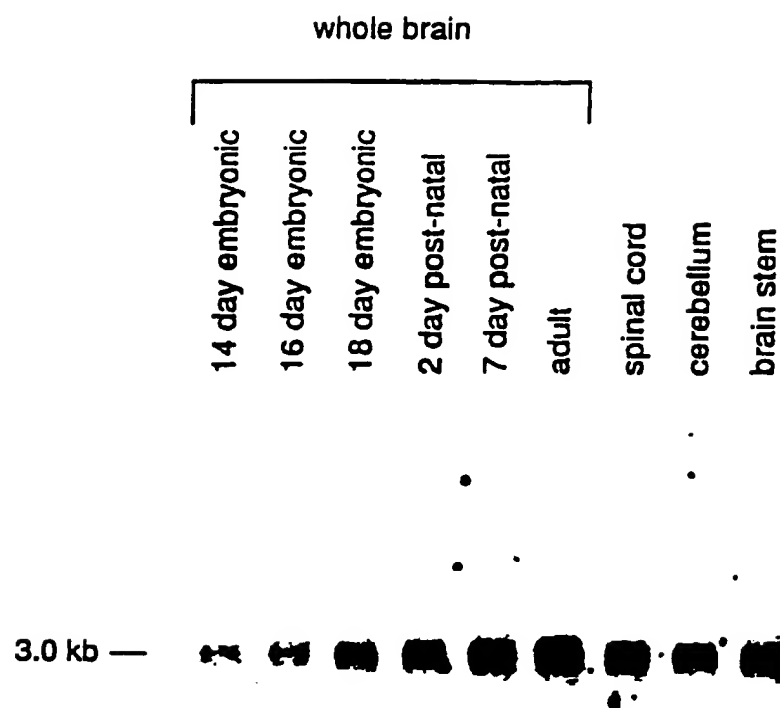
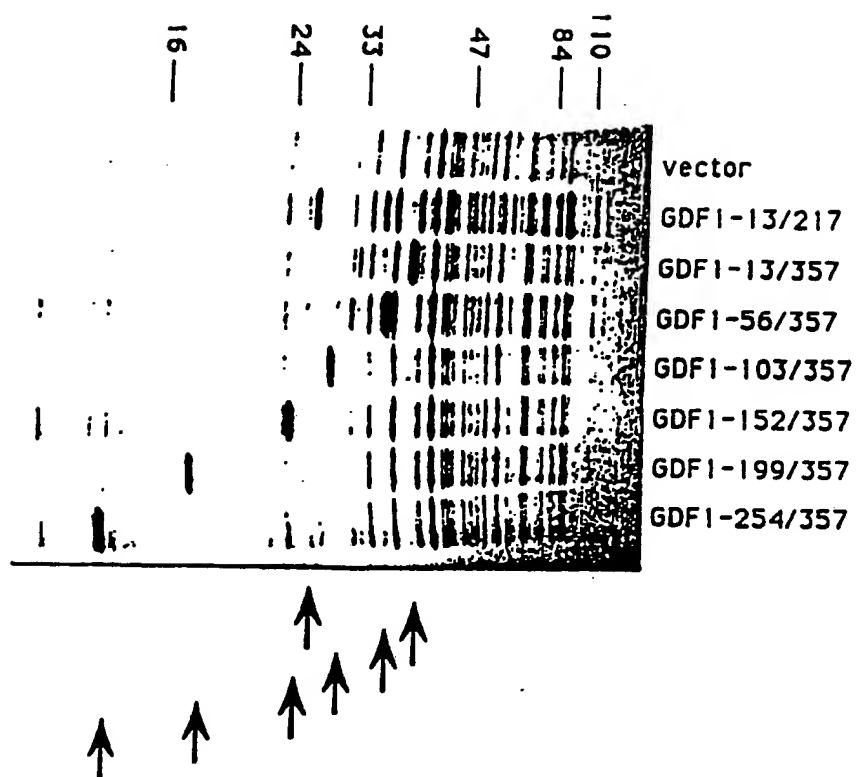


FIG. 8



11/19

FIG. 9



12/19

FIG. 10

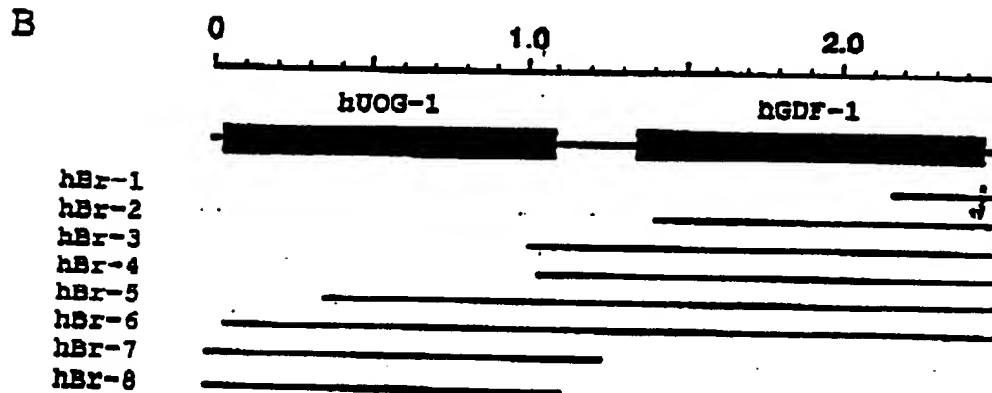
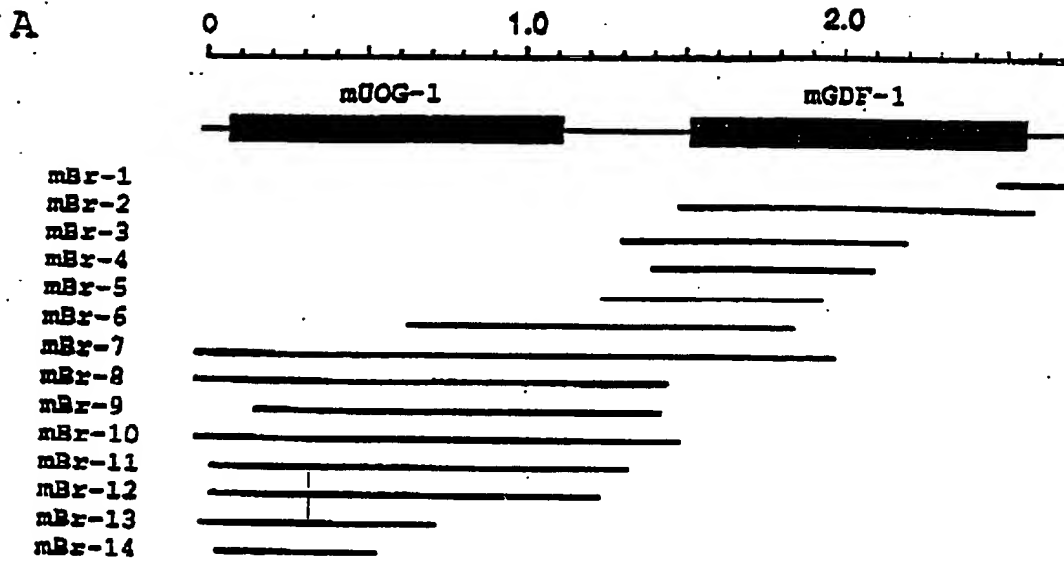


FIG. 11A

1 GCGCGTGACGCGAGGGCGCGCGGCGACTCGGACCGGTGCAGGCAACAGCGGAGACAGCGG 60
 61 AGAATTGGATAGCATGGCTGCTGCCGCGGCGACCCAGGCTCGAGGCGCCAGAGCCCAT 120
 M A A A A A T P R L E A P E P M
 121 GCCGAGTTATGCCGAGATGTTGCAACGAAGCTGGGCGCTCGGCGCTGGCGGCGGCTCAGGG 180
 P S Y A Q M L O R S W A S A L A A A Q G
 181 CTGCGGGGACTGCGGCTGGGACTGGCGCGCGCGGCTGGCGGAGCAGCGCACCTGGC 240
 C G D C G W G L A R R G L A E H A H L A
 241 TGCACCCGAGCTGCTGCTGGCGGTGCTGCGCTCTGGGGTGGACAGCGTGGCGTGGC 300
 A P E L L L A V L C A L G W T A L R W A
 301 AGCCACCACACATCTTTGCGCCCTGGCCAAGCGGTGCTGCTGCGAGCTAGAGATGC 360
 A T T H I F R P L A K R C R L Q P R D A
 361 TGCCAGGTTACCTGAGAGCGCTGGAAGCTTCTGTTCTACTTGGCCTGTTGGAGCTACTG 420
 A R L P E S A W K L L F Y L A C W S Y C
 421 CGCTTACCTGCTCCTGGGACCAAGTTATCCTTCTTCCATGACCGCGCTCTGCTCTCTA 480
 A Y L L L L G T S Y P F F H D P P S V F Y
 481 TGACTGGAGGTGAGGCTGCGAGTGCCTGGGACATCGCGGTGGCTATTGCTGCGAGG 540
 D W R S G M A V P W D I A V A Y L L Q G
 541 GAGTTTCTACTGCCACTCCATCTATGCCACCGTGTACATGGACAGCTGGCGTAAGGACTC 600
 S F Y C H S I Y A T V Y M D S W R K D S
 601 GGTGGTCACTGGTGCATCAGTGGTCAACCTGCTCCTATTGCTCTTCTACGCCCTT 660
 V V M L V H H V V T L L L I A S S Y A F
 661 CCGGTACCACAACTAGGCTCCTCGTGTCTTCTGCTGATGACGTGAGCGATGTGCAGCT 720
 R Y H N V G L L V F F L H D V S D V Q L
 721 GGAGTTCAAAAACCTCAACATCTACTTTAAGGCTAGGGGTGGTGCCTACCCTCGCTTGCA 780
 E F T K L N I Y F K A R G G A Y H R L H
 781 TGGGCTGGTGGCAACCTGGGCTGCTCAGCTTCTGTTTCTGCTGGTCTGGTTCGGCT 840
 G L V A N L G C L S F C F C W F W F R L
 841 CTACTGGTTCGCGCTCAAGGTTCTGTACGCCACTTGCCACTGCAGCTGCGAGTCTGTGCC 900
 Y W F P L K V L Y A T C H C S L Q S V P
 901 TGACATTCCGTACTACTTCTTCAACATTCTGCTGTGTCTGCTGATGGTCAATGACAT 960
 D I P Y Y F F F N I L L L L L M V M N I
 961 CTATTGGTTCCTGTACATTGTGGCTTTCGAGCCAAGGTGCTGACTGGTGCAGATGCGTGA 1020
 Y W F L Y I V A F A A K V L T G Q M R E
 1021 ACTGGAAGACTTGAGGGAGTACGACACTCTGGAAGCTCAGACAGCAAGCCCTCAAAAGC 1080
 L E D L R E Y D T L E A Q T A K P C K A
 1081 CGAGAAAGCACTGAGGAATGGCTGGTGAAGGACAAGCTCTTCTGAGTCTCTTGTCTCA 1140
 E K P L R N G L V K D K L F
 1141 ACTTCAGCCATCCAGGACTCTATCCCATCTACCTGGGATACTGACTCCGCCCTGGAGA 1200
 CTCGACCCAGTCCCTGGAGGTCTGCTCCACCCCTGGAGGCGCGGCTCCGCCCTTTGGCGG 1260
 1261 CATGGCTCGCCCTAGGACAATAGCCCCGCTTAAGATTGAGGATGCTACCCCTTCTCCA 1320
 1321 GGGACTCTGGCTGCCAGCAGTCCGCCCTTTCAGATCAATTCTCGACCCACCTTGGGA 1380
 1381 CTGCCGCCAGTCTGCGCTCTGGATCAGTGGGGTCCAGACACGCCCTCCAGGACCTC 1440
 1441 AAAGCACCCCGACTAAGGTCAACAGCCACTGGCCCCAGACGCACTGGGCTCCGCTGA 1500
 1501 CTCTCTGGACACCTCTGGGAGGAAAATGCTCCCTGTCTGCCATCGTTTTTGGGACAC 1560
 M L P V C H R F C D H
 1561 CTCCTCTCTGCTCTTGTGCTGCCCTCGACGACCTGGCCCCCGGCCAGCATCCATGGGC 1620
 L L L L L L L P S T T L A P A P A S H G
 1621 CCCGCTGCCGCTGCTCCAGGTTCTTGGGCTTCCGGAAGCGCCCGGAGCGTCCCCACA 1680
 P A A A L L Q V L G L P E A P R S V P T
 1681 CACCGACCTGTGCTCTGTATGTGGCGCTATTCCGTGCGCGGACCCCGAGGAGGCC 1740
 H R P V P P V M W R L F R R R D P Q E A
 1741 AGAGTGGGACGCGCTCTGCGGCCATGCCAGCTGGAGGAACTAGGGGTGCGCGGAAACAT 1800
 R V G R P L R P C H V E E L G V A G N I
 1801 GTGCGCCACATCCCCGACAGCGGTCTGTCTCCAGGCGCGCACAAACCCCGAGGACCTCG 1860
 V R H I P D S G L S S R P A Q P A R T S
 1861 GGGCTGTGCCCGAGTGGACAGTCTGTTGACTGTGCAATGTGGAGCCACAGAGCGC 1920
 G L C P E W T V V F D L S N V E P T R
 1921 CCAACACGCGCGCTTAGAGTTGCGGCTGGAGGCTGAGAGTGAAGATACAGGGGGTGG 1980
 P T R A R L E L R L E A E S E D T G G W
 1981 GAGCTAAGCGTGGCACTGTGGGCGGACGAGCATCCAGGGCTGAGTGTGCGCGGTG 2040
 E L S V A L W A D A E H P G P E L L R V
 2041 CCGGCGCCACAGGGGTGCTCCTGCGCGCAGACCTACTGGGACTGCAGTAGCCGCCAAC 2100
 P A P P G V L L R A D L L G T A V A N
 2101 GCATCAGTGCCCTGTACTGTGCGCTGGCGCTGTCACTGCACCTGGGGCCACTGCAGCC 2160
 A S V P C T V R L A L S L H . P G A T A A
 2161 TGTGGGCGCTGGCTGAGGCTCCCTGCTGCTGGTGACGCTGGACCCACGCTGTGTCCC 2220
 C G R L A E A S L L L V T L D P R L C P
 2221 TTGCCGCGATTGCGGCGCCACAGGAGCCAGGGTAGAAGTTGGTCCAGTGGGCACTTGT 2280
 L P R L R R R H T E P R V E V G P V G T C
 2281 CGTACCCGACGGTTCATGTGAGCTTCGCTGAGGTGGGCTGGCAGCGTTGGGTGATCGCG 2340
 R T R R L H V S F R E V G W H R W V I A
 2341 CCGCGTGGCTTCTAGCAACTTCTGCCAGGGCAGTGGCGACTACCCGAAACGCTGAGG 2400
 F R G F L A N F C Q G T C A L P E T L R
 2401 GGACCCGCGGCGCGCTGCACTCAACCACGCTGTGCTGCGCGGCTCATGCACGAGCT 2460
 G P G P P A L N H A V L R A L M H A A
 2461 GCTCCACCCCGGGTGCAGGCTCGCCCTGCTGCGTGCAGAGCGTCTATCACCCATCTCC 2520
 A P T P G A G S P C C V P E R L S P I S
 2521 GTGCTCTTCTGCAATAGTGACAACGTGGTCTGCGACACTACGAAGCATGGTGGTG 2580
 V L F F D N S D N V V L R H Y E D M V V
 2581 GATGAGTGTGGTGGCTTGACACCCGGGACACCTTTAGGGACCGCCCCACGCAAAA 2640
 D E C G C R
 2641 GCAGGACTGTTTGTTCATGTTTTATTGGTGACAAAAAGCTTAAACAAATTGACTAAA 2700
 2701 AATTAAAGTTCC 2711

14 / 19

FIG. 11B

1 GGACACGGCGGGCGAGCGGGCGGTATGGCGGGCGGGGGCCCGCGGGGGCCGACGGGG 60
M A A A G P A A G P T G
61 CCCGAGCCCATGCCGAGCTACCGCAGCTAGTGCAGCGCGGCTGGGGCAGCGCGCTGGCG 120
P E P M P S Y A Q L V Q R G W G S A L A
121 GCGGGCGGGGCTGCACGACTGCGGCTGGGGGCTGGCGGCTCGCGGCTGGCTGAGCAC 180
A A R G C T D C G W G L A R R G L A E H
181 GCGCACTGGCGCGCCCGAGCTGCTGCTGCTGGCGCTCGGGCGGCTGGGCTGGACCGCG 240
A H L A P P E L L L L L A L G A L G W T A
241 CTGGCTCCGCGGCCACTGCGGCGCTTTTCGGGCCCTGGCGAAGCGGTGCTGCCCTCCAG 300
L R S A A T A R L F R P L A K R C C L Q
301 CCCAGAGATGCCGCAAGATGCCGAGAGCGCTTGAAGTTTCTTCTACCTGGGCGAGC 360
P R D A A K M P E S A W K F L F Y L G S
361 TGGAGCTACAGTGCTACCTGCTGTTTGGCAGGACTACCCCTTCTTCCATGACCCACCA 420
W S Y S A Y L L F G T D Y P F F H D P P
421 TCTGCTTCTACGACTGGACGCGGGCATGGCAGTGCCACGGGACATTGACGCGGCTTAC 480
S V F Y D W T P G M A V P R D I A A A Y
481 CTGCTCCAGGGAAGCTTCTATGGCCACTCCATCTACGCTATACGCTATACGACCTGG 540
L L Q G S F Y G H S I Y A T L Y M D T W
541 CGCAAGGACTCGGTGGTCTGCTGCTCCACGAGTGGTCACTCTCATCCTCATCTGCTCC 600
R K D S V V M L L H H V V T L I L I V S
601 TCCTACGCTTCCGGTACCACAATGTGGGCTCCTTGTGCTCTTCTGACGATATCAGT 660
S Y A F R Y H N V G I L V L F L H D I S
661 GACGTGACGCTTGAATTCACCAAGCTCAACATTACTTCAAGTCCCGCGGCGCTCCTAC 720
D V Q L E F T T K L N I Y F K S R G G S Y
721 CATCGGCTGCATGCTTGGCAGCAGACTGGGCTGCTCAGCTTCCGCTTCACTGCTGTT 780
H R L H A L A A D L G C L S F G F S W F
781 TGGTTCGCGCTTACTGGTTCGCGCTCAAGGCTCCTGTATGCCACGAGTCACTGACGCTG 840
W F R L Y W F P L K V L Y A T S H C S L
841 CGCAGGCTGCTGACATCCCTTCTACTTCTTCAATGCGCTCCTGCTGCTCACC 900
R T V P D I P F Y F F F N A L L L L L T
901 CTATGAACCTTACTGGTTCCTGTACATCGTGGCGTTTGCAGCCAAGGTGTTGACAGGC 960
L M N L Y W F L Y I V A F A A K V L T G
961 CAGGTGCAGGCTGAAGGACCTGCGGAGTATGACACAGCGGAGGCCAGAGCCTGAAG 1020
Q V H E L K D L R E Y D T A E A Q S L K
1021 CCCAGCAAGCCGAGAAGCACTGAGGAACGGCCTGGTGAAGGACAAGCGCTTCTGAACC 1080
P S K A E K P L R N G L V K D K R F
1081 CCTCGGCCCCCGCCCGTGGACCGGCCCCCGCAATACCCGGGCCACGCTCCCGTGC 1140
CTTGGCCGCCCCCTCCACCCCTCAACTCTGCTCCTTAGGGCCGCGCCACCTCCCGTG 1200
1201 GGACCCGCGCCCTCATCTGCTTCAATTTCCGCGCCAGCCCCCAGGACCCCTGCCCC 1260
1261 TCCGGGGACACCGGCCCCCGCTCAGCCCACTGGTCCCGGGCGCGCGGACCCCTGCGCA 1320
1321 CTCTCTGGTTCATCGCTGGGAGGAAGATGCCACCGCCGAGCAAGGTCCCTGCGGCCACC 1380
M P P P Q Q G P C G H H
1381 ACCTCTCTCTCTCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
L L L L L L A L L P S L P L T R A P V P
1441 CCCAGGCCCAGCGCGCCCTGCTCCAGGCTTAGGACTGCGGATGAGCCCGAGGCTG 1500
P G P A A A L L Q A L G L R D E P Q G A
1501 CCCAGGCTCCGCGCGGTTCCCGGCTCATGTGGCGCTGTTTCGAGCGCGGACCCCC 1560
P R L R P V P P V M W R L F R R R D P Q
1561 AGGAGACCAGGTCTGGCTCGCGCGGAGCTCCCGAGGGTCAACCTGCAACCGTGCCACG 1620
E T R S G S R R T S P G V T L Q P C H V
1621 TGGAGGAGCTGGGGTGGCGGAACATCGTGGCCACATCCCGAGCGGCTGCGGCCA 1680
E E L G V A G N I V R H I P D R G A P T
1681 CCCGGGCTCGGAGCTGTCTCGGCGCGGGCAITGCCCTGAGTGGACAGTCTGCTTCG 1740
R A S E P V S A A G H C P E W T V V F D
1741 ACCTGTGCGCTGTGGAAACCGCTGAGCGCCGAGCGCGGCGCGCTGGAGCTCGGTTTCG 1800
L S A V E P A E R P S R A R L E L R F A
1801 CGGCGCGCGCGCGGAGCCCGAGGGCGGCTGGGAGCTGAGCGTGGCGCAAGCGGGCC 1860
A A A A A A P E G G W E L S V A Q A G Q
1861 AGGCGCGCGCGCGGACCCCGGGCGGCTGCTGCTCCGCAAGTGGTGGCGCCCTGGGGC 1920
G A G A D P G P V L L R Q L V P A L G P
1921 CGCCAGTGGCGCGGAGCTGCTGGGCGCGGCTTGGGCTGCAACCGCTCATGGCCGCGCA 1980
P V R A E L L G A A W A R N A S W P R S
1981 GCTCCGCTGGCGCTGGCGCTACGCCCCCGGGCCCTGCCGCTGCGCGGCTGGCGG 2040
L R L A L A L R P R A P A A C A R L A E
2041 AGGCTCGCTGCTGCTGCTGACCTCGACCGCGCTGTGCCACCCCTGGCCCCGGCGC 2100
A S L L L V T L D P R L C H P L A R P R
2101 GCGCGACGCGCAACCGTGTGGGGCGGGCGGCGGCGGCTGTGCGCGCGCGGGCG 2160
R D A E P V L G G G F G G A C R A R R L
2161 TGTACGTGAGCTTCCGCGAGGTGGGCTGGCAGCGCTGGGTCTCGCGCGCGCGGCTTCC 2220
Y V S F R E V G W B R W V I A P R G F L
2221 TGGCAACTACTGCCAGGTCAGTCCGCGCTGCCGCTGCGGCTGTGCGGGTCCGGGGGGC 2280
A N Y C Q G Q C A L P V A L S G S G G P
2281 CGCCGCGCTCAACACGCTGTGCTGCGCGGCTCATGCACGCGCGCGCGCGCGGAGCCG 2340
P A L N H A V L R A L M H A A A P G A A
2341 CGACCTGCGCTGCTGCTGCGCGCGGCTGTGCGCCACTCTCGGTGCTCTTCTTGACA 2400
D L P C C V P A R L S P I S V L F D N
2401 ACAGCGAACGTTGGTGGCGGAGTATGAGGACATGGTGGTGGACGAGTGGCGGCTGCC 2460
S D N V V L R Q Y E D M V V D E C G C R
2461 GCTAACCCGGGCGGGCAGGAGCGGGCCCAACAATAATGCCGCGTGG 2510

FIG. 12

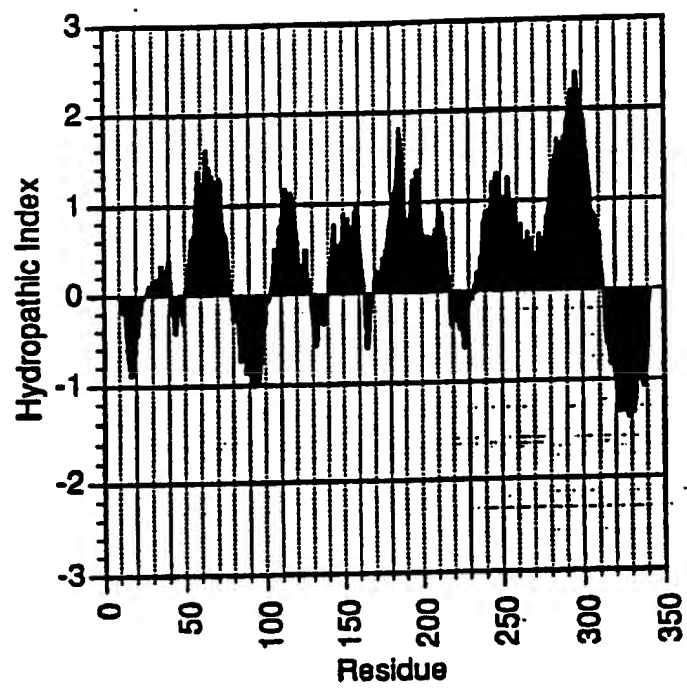


FIG. 13 A

	1		57
mGDF-1	MLPVCHRFCDH--LLLL-LLLFSTTLAFAPASMGFAAALLQVLGLPEAFRSVPTHRVPVP		
hGDF-1	MPPFQQGPGCGHLLLLLLALLLESPLTRAPVVPFGPAAALLQALGLRDEPQGAPRLRPVVP		
	1		60
	58		110
mGDF-1	VMWRLFRRRDPQEARVG-R---P---LRPCHVEELGVAGNIVRHIPDSGLSSRPAQPART		
hGDF-1	VMWRLFRRRDPQETRSGSRRTSPGVTLOPCHVEELGVAGNIVRHIPDRGAPTRASEPVSA		
	61		120
	111		166
mGDF-1	SGLCPewTVVFDLSNVEPTERETRARLELRLEAFCEDT--GGWELSVALWAD-AE-HPGP		
hGDF-1	AGHCFewTVVFDLSAVEPAERP'SRARLELRFAAAAAAAPEGGWELSVAQAGQGAGADPGP		
	121		180
	167		225
mGDF-1	ELLRVPAFP-GVLLRADLLGTAVAANASVPCTVRLALSLHPGATAACGRLAESLLLVTI		
hGDF-1	VLLRQLVPALGPPVRAELLGAAWARNASWERSRLALALRPAPAAACARLAESLLLVTI		
	181		240
	226		284
mGDF-1	DPRLC-PLERLRHTEPRVEVGFGVGTIRTRRLHVSFREVGVHWRWVIAPRGFLANFQGT		
hGDF-1	DPRLCHPLARFRDAEPVLGGGPGGARARRLYVSFREVGVHWRWVIAPRGFLANYQGC		
	241		300
	285		344
mGDF-1	ALPETLRGPGGPPALNHAIVLRALMHAAAPTFCAGSPVPERLSPISVLFFDNSDNVVLRL		
hGDF-1	ALFVALSGSGGPPALNHAIVLRALMHAAAPCA-ADLPVPARLSPISVLFFDNSDNVVLRL		
	301		359
	345	357	
mGDF-1	HYEDMVVDEICGR		
hGDF-1	QYEDMVVDEICGR		
	360	372	

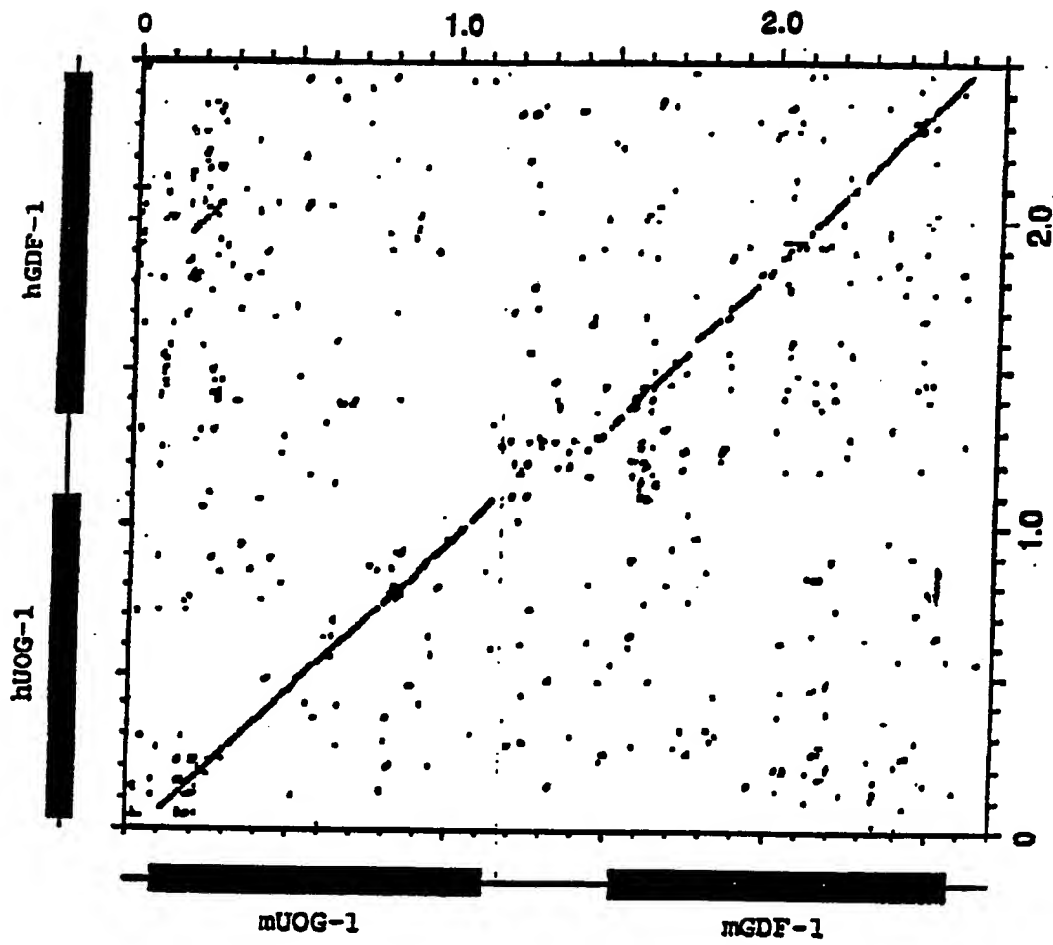
17/19

FIG. 13 B

	1	60
mUOG-1	MAAAAATFRLEAPEPMPSYAQMLORSWASALAAQCGDCGWGLARRGLAEHAHLAAPEL	
hUOG-1	MAAAGPAAGPTGPEPMPSYAQLVORGWSALAAARGCTDCGWGLARRGLAEHAHLAPPEL	
	1	60
	61	120
mUOG-1	LLAVLCALGWTALRWAATTHIFRPLAKRCRLQPRDAARLPESAWKLLFYLCWSYCAVLL	
hUOG-1	LLLALGALGWTALRSAATARLFRPLAKRCCLOPRDAAKMPESAWKFLFYLGWSYSAYLL	
	61	120
	121	180
mUOG-1	LGTSYPFFHDPPSVFYDWRSGMAVPWDIAVAYLLOGSFYCHSIYATVYMDSWRKDSVVML	
hUOG-1	SGTDYPFFHDPPSVFYDWTIPGMAVPRDIAAAYLLOGSFYGHSIYATLYMDTWRKDSVVML	
	121	180
	181	240
mUOG-1	VHHVVTLLLIASSYAFRYHNVGLLVEFLHDVSDVQLEFTKLNIFYKARGGAYHRLHGLVA	
hUOG-1	LHHVVTLLILIVSSYAFRYHNVGILVFLHDISDVQLEFTKLNIFYKSRGGSYHRLHALAA	
	181	240
	241	300
mUOG-1	NLGCLSF CFCWFWRFLYWFFLKVLYATCHCSLQSVDPDIPYFFFNILLLLLMVMNIYWFL	
hUOG-1	DLGCLSEGF SFWWRFLYWFFLKVLYATSHCSLRTVPDIPYFFFNALLLLLTLMNLYWFL	
	241	300
	301	350
mUOG-1	YIVAFAAKVL TGOMRELEDLREYDTLEAQ TAKPCKAEKPLRNGLVKDKLF	
hUOG-1	YIVAFAAKVL TGQVHELKDLREYDTAEASLKP SKAEKPLRNGLVKDKRF	
	301	350

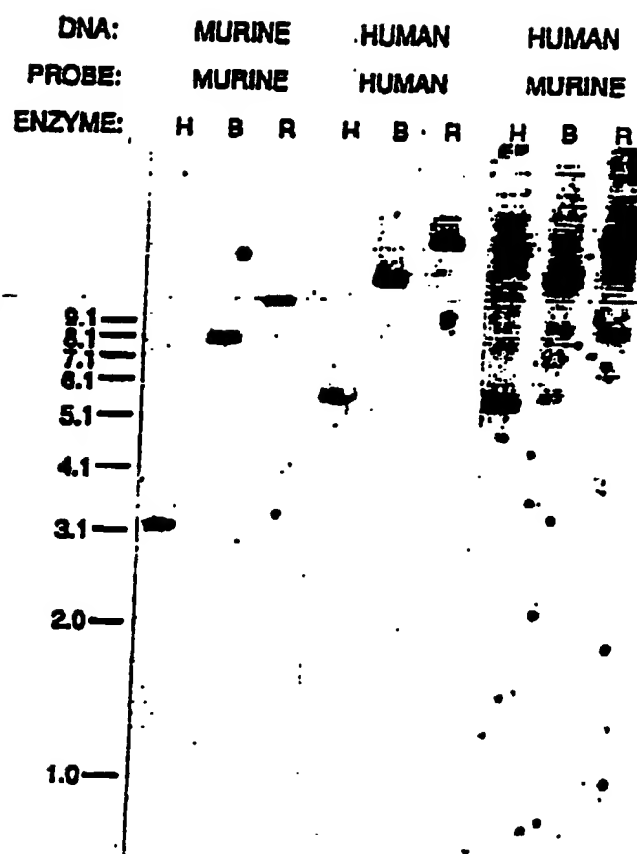
18/19

FIG. 13 C



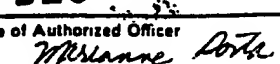
19/19

FIG. 14



INTERNATIONAL SEARCH REPORT

International Application No. PCT/US91/04096

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) ⁶			
According to International Patent Classification (IPC) or to both National Classification and IPC			
IPC(5): C12P 21/00, 21/02; C12N 15/00, 5/00			
U.S. Cl.: 435/69.4, 69.9, 172.1, 172.3, 240.1, 240.2, 320.1, 252.33; 536/27; 530/350, 399			
II. FIELDS SEARCHED			
Minimum Documentation Searched ⁷			
Classification System	Classification Symbols		
U.S.Cl.:	435/69.4, 69.9, 172.1, 172.2, 172.3, 240.1, 240.2, 320.1, 252.33; 536/27; 530/350, 399		
Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched ⁸			
APS, Dialog (files 5, 155, 351, 357 and 358), search terms: uog, gdf, tgf, transforming growth factor beta, superfamily, supergene, DNA, protein			
III. DOCUMENTS CONSIDERED TO BE RELEVANT ⁹			
Category ¹⁰	Citation of Document, ¹¹ with indication, where appropriate, of the relevant passages ¹²	Relevant to Claim No. ¹³	
Y, P	THE EMBO JOURNAL, Volume 9, No. 7, issued July 1990, E. Ozkaynak et al., "OP-1 cDNA Encodes an Osteogenic Protein in the TGF- β Family" pages 2085-2093, see entire document.	1-3, 1 1 - 1 4 , 1 6 , 19-20	
X Y	US, A, 4,886,747 (DERYNCK ET AL) 12 DECEMBER 1989, see entire document.	1, 3- 21 2	
X, P	Proceedings of the National Academy of Sciences, Vol. 88, issued May 1991, Lee, "Expression of growth/ differentiation factor 1 in the nervous system: Conservation of a bicistronic structure", pages 4250-4254, see entire document.	1-21	
Y	Proceedings of the National Academy of Sciences, Vol. 86, issued June 1989, Lyons et al., "Vgr-1, a mammalian gene related to <u>Xenopus VG-1</u> , is a member of the transforming growth factor β gene superfamily", pages 4554-4558, see entire document.	1-3, 1 1 - 1 4 , 1 6 , 17-20	
¹⁰ Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed		"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. "&" document member of the same patent family	
IV. CERTIFICATION			
Date of the Actual Completion of the International Search		Date of Mailing of this International Search Report	
12 November 1991		10 DEC 1991	
International Searching Authority		Signature of Authorized Officer	
ISA/US		 Marianne Porta	